

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

Philosophy of Artificial Intelligence
Russian Interdisciplinary Conference
in Honor of the 60-th Anniversary
of Artificial Intelligence Studies
March 17-18, 2016, Faculty of
Philosophy, Lomonosov Moscow State
University, Moscow

人工智能哲学

人工智能的哲学。
人工智能研究的六十周年
念全俄跨学科会
在2016年3月17-18日
莫斯科国立大学哲学系, 莫斯科

В Сборнике научных трудов представлены доклады и статьи участников Всероссийской междисциплинарной конференции «Философия искусственного интеллекта», которая состоялась 17-18 марта 2016 года на философском факультете Московского государственного университета имени М.В.Ломоносова. Конференция была организована Научным советом РАН по методологии искусственного интеллекта и когнитивных исследований и собрала ведущих ученых и специалистов данной проблематики.

Конференция проводилась при поддержке Российского фонда фундаментальных исследований (РФФИ), проект № 16-06-20063 Г.



МОСКВА • ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА • 2016



17-18 МАРТА 2016 г., ФИЛОСОФСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В.ЛОМОНОСОВА, г. МОСКВА

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ВСЕРОССИЙСКАЯ МЕЖДИСЦИПЛИНАРНАЯ КОНФЕРЕНЦИЯ

СБОРНИК НАУЧНЫХ ТРУДОВ

Посвящается шестидесятилетию исследований
искусственного интеллекта

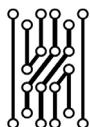
Москва, 2016

Научный совет РАН
по методологии искусственного интеллекта и
когнитивных исследований
Московский государственный университет имени М. В. Ломоносова,
философский факультет
Институт философии РАН

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА 2016

НАУЧНЫЕ ТРУДЫ
Всероссийской междисциплинарной конференции,
посвященной шестидесятилетию исследований
искусственного интеллекта, 17-18 марта 2016 г.,
философский факультет МГУ имени
М. В. Ломоносова, г. Москва

Под редакцией академика РАН, д.ф.н., проф. В. А. Лекторского,
д.ф.н., проф. Д. И. Дубровского, д.ф.н. А. Ю. Алексеева



ИИНТЕП

Москва
2017

УДК 100.32

СЕРИЯ

ББК 32.813 «ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

Ф 56

Редакционная коллегия:

В. А. Лекторский – академик РАН, д.ф.н., проф., гл.ред.,

А. Ю. Алексеев – д.ф.н., зам.гл.ред.,

Д. И. Дубровский – д.ф.н., проф., зам.гл.ред.,

С. Н. Васильев – академик РАН, д.ф.-м.н., проф.,

В. Л. Макаров – академик РАН, д.ф.-м.н., проф.,

В. С. Степин – академик РАН, д.ф.н., проф.,

К. В. Анохин – чл.-корр. РАН, д.мед.н., проф.,

Б. М. Величковский – чл.-корр. РАН, д.психол.н., проф.,

В. А. Глазунов – чл.-корр. РАН, д.ф.н., д.т.н., проф.,

А. М. Иваницкий – чл.-корр. РАН, д.мед.н., проф.,

В. В. Миронов – член-корр. РАН, д.ф.н. проф.,

Б. Г. Юдин – чл.-корр. РАН, д.ф.н., проф.,

В. И. Аршинов – д.ф.н., проф., *И. А. Бугаков* – д.т.н., проф.,

В. Г. Кузнецов, д.ф.н., проф., *О. П. Кузнецов* – д.т.н., проф.,

Ю. Ю. Петрунин – д.ф.н., проф.,

Т. В. Черниговская – д.б.н., д.филол.н., проф.,

В. К. Финн – д.т.н., проф., *Е. А. Янковская* – к.ф.н., *К. К. Кузичева*.

Ф 56 Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта, 17-18 марта 2016 г., философский факультет МГУ имени М. В. Ломоносова, г. Москва / Под ред. В. А. Лекторского, Д. И. Дубровского, А. Ю. Алексеева. – М.: ИИнтелл, 2017. – 340 с. – (Сер.: Философия искусственного интеллекта). – ISBN 978-5-98956-012-7.

Издание осуществлено при финансовой поддержке Российского фонда фундаментальных исследований, проект № 16-06-20063-г.

© НСМИИ РАН, 2016

© Авторы, 2016

© ООО «ИИнтелЛ», 2017

ISBN 978-5-98956-012-7

ОГЛАВЛЕНИЕ

Предисловие	9
Садовничий В. А. Искусственный интеллект и суперкомпьютер	17
Миронов В. В. Трансформация культур: от классической к электронной.....	28
Лекторский В. А. Философия, искусственный интеллект, когнитивные исследования	37
Дубровский Д. И. Сознание, мозг, искусственный интеллект: новые проблемы.....	45
Макаров В. Л. Искусственные общества	55
Степин В. С. Постнеклассическая рациональность и информационное общество	59
Анохин К. В. Когнитивные вычисления на основе нейронных гиперсетей.....	70
Юдин Б. Г. «Улучшение» человека и искусственный интеллект	87
Черниговская Т. В. Фуэте, фонема, формула, фотон: языки мозга и культуры	95
Глазунов В. А. Возможности применения роботов параллельной структуры в медицине.....	101
Кузнецов В. Г. Парадигмальный анализ теории искусственного интеллекта	108
Кузнецов О. П. Искусственный интеллект и науки о мозге – методологические проблемы.....	128
Смирнова Н. М. Философия искусственного интеллекта и феноменология: взаимодополнительность исследовательских стратегий	136
Алексеев А. Ю. Тест Тьюринга: 66 лет спустя.....	154

Кувакин В. А.

Я-мировоззрение: проект модификации структуры
внутреннего мира..... 189

Петрунин Ю. Ю.

Искусственный интеллект: наука, миф, философия..... 202

Мелик-Гайказян И. В.

Искусственный интеллект: семиотическая диагностика
аттракторов конвергентных технологий 208

Васюков В. Л.

«Витгенштейнианский» интенциональный искусственный
интеллект 218

Анисов А. М.

Компьютерные модели времени 236

Никитина Е. А.

Философско-методологические проблемы развития
интеллектуальной робототехники..... 257

Ершова Р. В.

Искусственный интеллект как психологический вызов 268

Шестакова М. А.

Методологические проблемы многоагентных систем..... 275

Михайлов И. Ф.

Искусственный интеллект и когнитивные науки: перспектива
антирепрезентационализма 284

Янковская Е. А.

Гетерархическая модель в рамках коннекционизма..... 295

Катречко С. Л.

О соотношении искусственного и естественного интеллекта..... 304

Дискуссия 320

Заключение..... 335

CONTENT

Introduction.....	9
<i>Sadovnichy, V. A.</i> Artificial Intelligence and Supercomputer.....	17
<i>Mironov, V. V.</i> Transformation of cultures: from classical to electronic.....	28
<i>Lektorsky, V. A.</i> Philosophy, artificial intelligence, cognitive research.....	37
<i>Dubrovsky, D. I.</i> Consciousness, the brain, artificial intelligence: new problems.....	45
<i>Makarov, V. L.</i> Artificial societies.....	55
<i>Stepin, V. S.</i> Post-nonclassical rationality and the information society.....	59
<i>Anokhin, K. V.</i> Cognitive calculations based on neural hypernetworks.....	70
<i>Yudin, B. G.</i> «Improvement» of man and artificial intelligence	87
<i>Chernigovskaya, T. V.</i> Fute, phoneme, formula, photon: languages of the brain and of culture	95
<i>Glazunov, V. A.</i> The possibilities of using parallel robots in medicine.....	101
<i>Kuznetsov, V. G.</i> Paradigm analysis of the theory of artificial intelligence	108
<i>Kuznetsov, O. P.</i> Artificial intelligence and sciences of brain – methodological problems	128
<i>Smirnova, N. M.</i> Philosophy of artificial intelligence and phenomenology: the complementarity of research strategies.....	136
<i>Alekseev, A. Yu.</i> Turing test: 66 years later	154
<i>Kuvakin, V. A.</i> I-worldview: a project to modify the structure of the inner world.....	189

Petrinin, Yu. Yu.	
Artificial intelligence: science, myth, philosophy.....	202
Melik-Gaykazyan, I. V.	
Artificial intelligence: Semiotic Diagnostics of Convergent Technology Attractors.....	208
Vasiukov, V. L.	
Wittgensteinian Intentional Artificial Intelligence	218
Anisov, A. M.	
Computer models of time	236
Nikitina, E. A.	
Philosophical and methodological problems of the development of intelligent robotics.....	257
Ershova, R. V.	
Artificial intelligence as a psychological challenge.....	268
Shestakova, M. A.	
Methodological problems of multi-agent systems	275
Mikhailov, I. F.	
Artificial intelligence and cognitive science: antirepresentationalist perspective.....	284
Yankovskaya, E. A.	
The heterarchic model within the framework of connectivity	295
Katrechko, S. L.	
On the Relation between Natural and Artificial Intelligence	304
Discussion	320
Conclusion.....	335

目录

介绍	9
Sadovnichy, V. A. 萨多伏拟其《人工智能和超级计算机》.....	17
Mironov, V. V. 米落诺扶《文化转变：从古典时代到电子时代》.....	28
Lektorsky, V. A. 列可多斯及《哲学，人工智能，认知研究》.....	37
Dubrovsky, D. I. 读布落扶其《意识，脑，人工智能，新问题》.....	45
Makarov, V. L. 马卡落扶《人工社会》.....	55
Stepin, V. S. 斯切兵《合理性和信息社会》.....	59
Anokhin, K. V. 阿诺轻《基于神经超网络的认知计算》.....	70
Yudin, B. G. 幼丁《“人类改善”与人工智能》.....	87
Chernigovskaya, T. V. 车尔泥古扶斯卡亚, 扶矮特《音位，公式，光子： 脑子和文化语言》.....	95
Glazunov, V. A. 各拉祖诺扶《医学中使用平行结构机器人的能力》.....	101
Kuznetsov, V. G. 库姿妥册扶《人工知能理论的范式分析》.....	108
Kuznetsov, O. P. 库姿妥册扶《人工智能寻找新范式的理论》.....	128
Smirnova, N. M. 斯米尔诺罚《人工智能哲学与现象学：研究战略互补性》.....	136
Alekseev, A. Yu. 阿列写扶, 余《图灵(«Turing»)试验：66年后》.....	154
Kuvakin, V. A. 库瓦琴《我是世界观：心理结构的改造工程》.....	189
Petrinin, Yu. Yu. 备土露宁《人工智能：科学，神话，哲学》.....	202

Melik-Gaykazyan, I. V. 美理刻盖杂强 《人工智能：融合技术吸引子会聚的诊断》.....	208
Vasiukov, V. L. 瓦锈可副 《“维特根斯坦”意图人工智能》.....	218
Anisov, A. M. 阿泥所扶 《时间的电脑模型》.....	236
Nikitina, E. A. 泥其其拿 《智能机器人发展的哲学和方式论问题》.....	257
Ershova, R. V. 业尔说法 《人工智能作为一个心理办法》.....	268
Shestakova, M. A. 设寺它可瓦 《多代理系统的方式问题》.....	275
Mikhailov, I. F. 米骇落伏 《人工智能与认知科学：反展示性的远景》.....	284
Yankovskaya, E. A. 养可扶丝卡压 《教育中的知识系统》.....	295
Katrechko, S. L. 卡特类其可 《人工智能与大脑智能的相比》.....	304
讨论	320
结论	335

ПРЕДИСЛОВИЕ

INTRODUCTION

介绍

Насколько нам известно, нигде в мире, кроме нашей страны, не проводятся конференции, подобные конференции «Философия искусственного интеллекта» («Philosophy of Artificial Intelligence»), то есть специально посвященные философским проблемам искусственного интеллекта (ИИ). Конечно, организуются секции, круглые столы, симпозиумы, на которых изучаются методологические и мировоззренческие проблемы искусственного интеллекта. Однако они проводятся в составе других, более крупных мероприятий.

Аналогичные философские конференции были в Советском Союзе: в 1962 году прошла Первая конференция по методологическим вопросам кибернетики. Вторая конференция состоялась в 1970 году. Это был период всплеска научных достижений отечественной кибернетики. Затем интерес к философскому осмыслению этой проблематики стал снижаться, и весьма долгое время подобных конференций не проводилось.

Двенадцать лет назад мы возродили эти мероприятия и провели в январе 2005 года Первую всероссийскую междисциплинарную конференцию «Философия искусственного интеллекта» (руководители – академики РАН В.А. Лекторский, В.Л. Макаров и профессора Д.В. Быков, Д.И. Дубровский, Г.С. Осипов). Она состоялась в Московском государственном институте электроники и математики (этого института сейчас нет). В работе конференции приняло участие свыше трёхсот человек, в том числе крупные ученые из Армении, Беларуси, Болгарии, Вьетнама, Германии, Грузии, Украины, Франции. На конференции выступали крупные специалисты в области компьютерных наук, биологии, математики, медицины, психологии, техники, физики, филологии, философии, химии и экономики и др.

Острый интерес научной общественности, проявленный к философским проблемам искусственного интеллекта, чрезвычайно многообразный спектр проблем, междисциплинарный статус дискуссий, наконец, социокультурные реалии современного общества – все это послужило поводом для создания нашего Научного совета РАН по методологии искусственного интеллекта. Совет был учрежден в феврале 2005 года и получил статус научной общественной организации при Отделении общественных наук РАН, а 17 марта 2005 г. было проведено первое заседание Совета и первое заседание постоянно действующего междисциплинарного научно-теоретического семинара «Философско-методологические проблемы искусственного интеллекта», на котором с докладом «Сознание, мозг, искусственный интеллект» выступил проф. Д. И. Дубровский. К настоящему времени состоялось 83 заседания этого семинара.

Философские и методологические исследования искусственного интеллекта, во многом благодаря работе нашего Совета, стали систематическими (координатор научных программ д.ф.н. А. Ю. Алексеев). Ежегодно проводятся молодежные конференции «Искусственный интеллект: философия, методология, инновации» (рук.: академик РАН А. С. Сигов) Одна из этих конференций прошла на базе СПбГУ, все остальные – в МИРЭА (уч. секр.: д.ф.н. Н. А. Никитина), в этом году прошла десятая такая конференция.

В Институте философии РАН ежегодно организуются мероприятия, посвященные проблемам сознания, самосознания, субъективности, на которых обсуждается когнитивная тематика. На базе вузов и академических структур наших региональных отделений так же организуются крупные мероприятия. Например, в г. Иваново ежегодно проводилась конференция «Актуальные проблемы современной когнитивной науки» (рук.: проф. Т. Б. Кудряшова), в г. Пермь – «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века» (рук.: проф. Л. Н. Ясницкий).

На протяжении ряда лет работают секции, проблемы которых тесно сопряжены с философско-методологическими исследованиями искусственного интеллекта: «Нейрофилософия» (рук.: проф. В. Г. Кузнецов, философский факультет МГУ), «Управление знаниями» (рук.: проф. Ю. Ю. Петрунин, факультет государственного управления МГУ), «Проблемы информатики» (рук.: проф. К. К. Колин), «Электронная культура» (рук.: д.ф.н. А. Ю. Алексеев, философский факультет МГУ и Московский государственный институт культуры),

«Философия творчества» (рук. : проф. Н. М. Смирнова, Институт философии РАН), «Рефлексивные процессы и управление» (рук. : проф. В. Е. Лепский, Институт философии РАН), «Музыка, математика, естествознание» (рук. : проф. А. А. Кобляков, Московская государственная консерватория им. П. И. Чайковского). Систематически проводятся и другие мероприятия по философии и методологии когнитивных исследований.

Поэтому философская проблематика исследований искусственного интеллекта получает прочные научно-организационные основания и, следовательно, теоретические исследования и технологические проекты становятся рационально обоснованными и соразмерными с собственно человеческими интенциями и социальными императивами.

О конференции

Проблематика современных философских исследований искусственного интеллекта раскрывалась в докладах и выступлениях ведущих отечественных ученых как в рамках пленарной части конференции (17 марта 2016 г.), так и в ходе проведения широкой дискуссии (18 марта 2016 г.). Пленарным заседанием руководили академик РАН, заведующий сектором теории познания ИФ РАН В. А. Лекторский, декан философского факультета МГУ, чл.-корр. РАН В. В. Миронов и главный научный сотрудник ИФ РАН, проф. Д. И. Дубровский. Дискуссией руководили представители кафедры философии и методологии науки философского факультета МГУ им. М. В. Ломоносова: заведующий кафедрой, проф. В. Г. Кузнецов, в. н. с. А. Ю. Алексеев и доц. А. В. Чусов.

В Предисловии мы кратко охарактеризуем идеи пленарных докладов конференции. В Заключении будет представлен обзор докладов второго дня конференции.

Собственно философский вектор исследования проблем искусственного интеллекта был задан сразу в первых двух докладах: академика РАН, доктора физико-математических наук, профессора, ректора МГУ имени М. В. Ломоносова В. А. Садовниченко на тему «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СУПЕРКОМПЬЮТЕР» и члена-корреспондента РАН, доктора философских наук, декана философского факультета МГУ В. В. Миронова «ТРАНСФОРМАЦИЯ КУЛЬТУР: ОТ КЛАССИЧЕСКОЙ К ЭЛЕКТРОННОЙ». Как сопрягается сфера высокопроизводительных параллельных вычислений, с одной стороны, с параметрами человеческой, социокультурной деятельности?

В. А. Садовничий подчеркнул, что ИИ – это предмет междисциплинарных исследований предельно широкого диапазона. Над этим предметом работают философы, биологи, психологи, лингвисты, социологи, специалисты в области вычислительной техники, информатики, программирования, робототехники и, конечно, математики. Закономерно, что эти проблемы обсуждаются именно сейчас, так как назрело время, когда надо скоординировать, систематизировать, интегрировать эти различные направления действий ученых. И именно философам, гуманитариям самого широкого профиля надлежит это сделать, наметить перспективы конвергентных технологий и понять, что это означает для человечества. Исследования искусственного интеллекта начались, все-таки, с математики. Также в своем выступлении В. А. Садовничий отметил, что настоящей проблемой остается вопрос о том, как смоделировать реальные психические процессы, а не просто формализовать некоторые процедуры мышления. Он подчеркнул, что особую роль в моделировании интеллекта играет использование суперкомпьютерных технологий. Предоставляемые этими технологиями вычислительные мощности помогают решать многофакторные задачи, возникающие при исследовании интеллекта, которые невозможно решить с помощью обычного моделирования.

В докладе В. В. Миронова акцент был сделан на том, что искусственный интеллект – это явление, связанное не только с наукой, но и с культурными аспектами. Поэтому важно рассматривать социокультурные проблемы, связанные с развитием технологий. Становление новой, электронной культуры, совпадает с процессами глобализации, которые приводят к появлению транснационального рынка, необходимости интеграции различных государств в различных транснациональных системах, что осуществляется в условиях глобальной системы коммуникации, которая влияет на изменение диалогов между культурами. Деформируется характер диалога как семиосферы, и эта деформация оказывает влияние на компоненты самих культур. Соответственно, меняются и характеристики нашего мышления. В электронной культуре изменяется роль экспертизы и роль эксперта. Электронная культура предоставляет широчайшие возможности поиска в почти безграничном поле информации. В связи с этим экспертиза часто сводится к поиску необходимой информации. Это приводит к изменению эпистемологического характера знаний, так как мы уже затрудняемся оценить объективность этого поиска информации.

В докладе академика, доктора философских наук, заведующего сектором теории познания Института философии РАН В.А. Лекторского «ФИЛОСОФИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, КОГНИТИВНЫЕ ИССЛЕДОВАНИЯ» кратко рассматривалась история формирования научного направления «искусственный интеллект». Это направление возникло 60 лет назад. Вызов был сформулирован самой идеей искусственного разума о том, что машина может что-то делать лучше человека. Сегодня идея соотношения искусственного и естественного стала весьма острой. Сейчас выдвигается идея, что можно искусственно создать всё то, что было естественным, и то, что природа создать не смогла. Это уже вызов космосу. Человечество на новом этапе может превратиться в демиурга вселенной, управлять эволюцией. Возникают этические проблемы, связанные с разработками в области искусственного интеллекта. К их числу относится проблема свободы воли у искусственной интеллектуальной системы и вообще вопрос о необходимости разработки «сильного» искусственного интеллекта, который может негативно воздействовать на своих создателей.

В докладе сопредседателя НСМИИ РАН, академика, доктора физико-математических наук, профессора, директора Центрального экономико-математического института РАН В. Л. Макарова «ИСКУССТВЕННЫЕ ОБЩЕСТВА» рассматривается направление исследований, которое занимается компьютерным моделированием социальных процессов. Искусственные общества представляют собой многоагентные системы. Агенты обладают сложным поведением, их взаимодействие формирует определенные устойчивые структуры, которые можно обозначить как искусственные социальные системы. В искусственных обществах это чисто числовые компьютерные эксперименты. Возникает вопрос традиционной теории познания: имеет ли это отношение к действительности.

В докладе академика, доктора философских наук, профессора В. С. Степина «ПОСТНЕКЛАССИЧЕСКАЯ РАЦИОНАЛЬНОСТЬ И ИНФОРМАЦИОННОЕ ОБЩЕСТВО» рассматривалась проблематика искусственного интеллекта с точки зрения теории сложных систем. Системы искусственного интеллекта можно отнести к третьему типу систем или объектов — сложные саморазвивающиеся системы, переходящие от одного типа гомеостаза к более сложному типу гомеостаза. Концепция сложных развивающихся систем — то, что требует особого типа рациональности, рефлексии, понимания, объяснения,

описания и так далее, то есть новых идеалов и норм науки. В постнеклассическом подходе, в котором исследуются саморазвивающиеся системы, радикально меняется проблема естественного и искусственного. Природа не реализовала всего того, что могла бы реализовать. Но, создав человека, она запускает через него новые линии своей эволюции. Так решается эта проблема: искусственное и естественное неразличимы с точки зрения саморазвивающейся системы.

В докладе «ВОЗМОЖНОСТИ РОБОТОВ ПАРАЛЛЕЛЬНОЙ СТРУКТУРЫ» В.А. Глазунова — чл.-корр. РАН, д.ф.н., д.т.н., проф., директора Института машиноведения РАН имени А.А. Благоднарова — рассматривались примеры конструирования систем, которые явно не обладают интеллектуальными способностями, но могут производить вполне интеллектуальные манипуляции во внешнем мире. Благодаря параллельной схеме действий такой робот превосходит возможности биологических существ, включая человека: естественные организмы, как правило, способны к осуществлению только последовательных действий.

Д. И. Дубровский, сопредседатель НСММИ РАН, д.ф.н., проф., главный научный сотрудник Института философии РАН в докладе «СОЗНАНИЕ, МОЗГ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НОВЫЕ ПРОБЛЕМЫ» подвел промежуточные итоги разработки данной темы, которая была начата на конференции ФИИ-2005 г. Этой теме была посвящена работа отдельной секции. Докладчик подчеркнул, что исследование естественного интеллекта, психических процессов и феномена сознания — важнейшая задача современной науки и практики, источник новых подходов в разработке информационных и когнитивных технологий, а тем самым необходимое условие продуктивного развития всего комплекса НБИКС-конвергенции. Проблематика изучения сознания стала главным условием дальнейшего развития цивилизационных процессов.

Ряд других докладов был посвящен поиску методологических связей между символьным (репрезентативным, эвристическим, семиотическим) и коннекционистским (арепрезентативным, бионическим, нейрофизиологическим) подходом к ИИ.

Проф., д.ф.н., зав. кафедрой философии и методологии науки философского ф-та МГУ В. Г. Кузнецов в докладе «ПАРАДИГМАЛЬНЫЙ АНАЛИЗ ТЕОРИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» представил историю становления и развития символьного и коннекционистского подходов в теории и технических приложениях искусственного

интеллекта. Общие положения концепции совпадают с парадигмальной философией науки Томаса Куна, полагает докладчик.

Т. В. Черниговская, д.б.н., д.филол.н, проф., зав. лаб. когнитивных исследований и кафедрой проблем конвергенции естественных и гуманитарных наук СПбГУ, г. Санкт-Петербург, в докладе «ФУЭТЕ, ФОНЕМА, ФОРМУЛА, ФОТОН: ЯЗЫКИ МОЗГА И КУЛЬТУРЫ» показала возможность тесной взаимосвязи естественнонаучной и социогуманитарной методологии именно в контексте изучения мозговой активности в связи с психическими феноменами. В свою очередь, она говорила об алармистских настроениях, связанных с тем, что все люди, оказавшись в электронной культуре, неожиданно и вдруг почувствовали себя в ситуации пространственной, временной, онтологической и праксеологической неопределенности: в условиях компьютерных коммуникаций возникает вопрос «Что принадлежит мне, где, когда, кто я и способен ли я быть Я?».

К. В. Анохин, член-корр. РАН, д.мед.н., рук. отдела нейронаук НИЦ «Курчатовский институт», руководитель лаборатории нейробиологии памяти НИИ нормальной физиологии им. П. К. Анохина в докладе «КОГНИТИВНЫЕ ВЫЧИСЛЕНИЯ НА ОСНОВЕ НЕЙРОННЫХ ГИПЕРСЕТЕЙ» указывал на необходимость разрешения двух принципиальных проблем в области различия между компьютерами и естественным мозгом. Это достигается, по мнению докладчика, теорией когнитома. Лучший аппарат, который к настоящему моменту удалось найти для необходимой формализации когнитома – это теория гиперсетей. Биологическая или субстанцированная модель – это гиперсетевая теория мозга.

На возможность компьютерного моделирования нейронных ансамблей и, в связи с этим, на традиционные логико-эпистемологические и специфические проблемы программирования переходов от отдельных нейронов к их паттернам указал Председатель Научного совета Российской ассоциации искусственного интеллекта, д.т.н., проф., зав. лаб. Института проблем управления РАН О. П. Кузнецов в докладе «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПОИСКАХ НОВОЙ ПАРАДИГМЫ».

Б. Г. Юдин, член-корреспондент РАН, д.ф.н., проф., зав. отделом комплексных проблем изучения человека Института философии РАН, в докладе «"УЛУЧШЕНИЕ" ЧЕЛОВЕКА И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ» отметил ряд евгенических проблем биотехнологий в связи с развитием искусственного интеллекта.

Благодарности

Выражаем благодарность Российскому фонду фундаментальных исследований (РФФИ) за финансовую поддержку конференции «Философия искусственного интеллекта» как в 2005 году (проект № 05-06-85002-г), так и в этом году (проект № 16-06-20063-г).

За организационную помощь выражаем глубокую благодарность главному ученому секретарю Ученого совета МГУ имени М. В. Ломоносова, доктору филологических наук, профессору Ольге Владимировне Раевской и секретарю Центрального экономико-математического института РАН Марии Бурылиной.

Бакалавров философского факультета, выпускников 2017 года, Никиту Борисова, Елену Манжелей, Жанну Лихачеву, Викторину Бродскую и Степана Козлова благодарим за качественную и оперативную расшифровку стенограмм пленарных докладов. За дизайн конференции благодарим Тодора Пожарева, аспиранта кафедры философии и методологии науки философского факультета МГУ имени М. В. Ломоносова.

Профессор Сюэ-Фэн Юань (Xue-Feng Yuan) из университета г. Гуанчжоу (КНР), организатор первого рабочего совещания стран БРИКС по вопросам высокопроизводительных вычислений и информационным технологиям (апрель 2017 года, г. Гуанчжоу), подсказал идею перевода названий статей на китайский язык, а профессиональный переводчик Ангелина Коблякова проделала эту сложную работу.

Председатель организационного комитета конференции

В. А. Лекторский

Сопредседатели программного комитета конференции

Д. И. Дубровский

В. Г. Кузнецов

В. В. Миронов

Заместитель председателя организационного комитета

А. Ю. Алексеев

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СУПЕРКОМПЬЮТЕР ARTIFICIAL INTELLIGENCE AND SUPERCOMPUTER

萨多伏拟其 «人工智能和超级计算机

**Садовничий
Виктор Антонович**

академик РАН, доктор физико-математических наук, профессор, ректор Московского государственного университета им. М. В. Ломоносова, г. Москва

Sadovnichy Victor A.

academician of the Russian Academy of Sciences, Doctor of Science (Physical and Mathematical), professor, rector of the Lomonosov Moscow State University, Moscow



Аннотация. Обозначена роль ученых МГУ им. М. В. Ломоносова в развитии исследований искусственного интеллекта. Изучаются особенности применения суперкомпьютеров в моделировании мышления, создании нейроинтерфейсов, систем виртуальной реальности. В суперкомпьютерных технологиях повышается роль эвристического подхода к постановке и программированию математических задач. Дается краткая характеристика суперкомпьютера МГУ.

Ключевые слова: МГУ имени М. В. Ломоносова, искусственный интеллект, суперкомпьютер, виртуальная реальность, нейроинтерфейс, математические эвристики.

Annotation. The role of Lomonosov Moscow State University in the development of domestic and world studies of artificial intelligence. The peculiarities of the use of supercomputers in modeling thinking, creating neurointerfaces, virtual reality systems are studied. In supercomputer technologies, the role of the heuristic approach to the formulation and programming of mathematical problems increases. A brief description of the supercomputer of the Moscow State University is given.

Keywords: Lomonosov Moscow State University, artificial intelligence, supercomputer, virtual reality, neurointerface, mathematical heuristics.

Для меня большая честь выступать на конференции, посвященной философии. Это мой второй доклад, второе выступление, первое было несколько лет назад. Я делал доклад на тему «Наука как метафора».

Конечно, я – математик. Проблемы, которыми я сам занимаюсь, не являются простыми и их невозможно прямо связать с философскими понятиями, потому что они ближе к математике. Но с искусственным интеллектом эти проблемы связываются легко.

Искусственный интеллект – это предмет междисциплинарных исследований предельно широкого диапазона. Над этим предметом работают философы, биологи, психологи, лингвисты, социологи, специалисты в области вычислительной техники, информатики, программирования, робототехники и, конечно, математики. Закономерно, что эти проблемы стали сейчас обсуждаться. Назрело время, когда надо скоординировать, систематизировать, интегрировать эти различные направления действий ученых. И, наверное, именно философам, гуманитариям самого широкого профиля, надлежит это сделать, наметить перспективы конвергентных технологий и понять, что это означает для человечества.

Все, что связано с изучением человеческого мозга, – это, безусловно, приоритет современной науки. У всех ведущих держав есть свои проекты исследования мозга, они возведены в ранг национальных программ. В Европе работают над десятилетним мегапроектом «Человеческий мозг». Европейский Союз выделил более миллиарда долларов на создание компьютерной модели мозга человека. В 2013 году президент США Барак Обама выдвинул «brain-инициативу». По его словам, это вызов 21-го века. Начальное финансирование – более 100 млн. долларов. А Китай в свою программу вкладывает в несколько раз больше средств, чем Соединенные Штаты. И все это с амбициозной целью – сделать из мозга как величайшего чуда природы высокотехнологичный артефакт. Как сказал нобелевский лауреат Джеймс Уотсон: «Мозг – это последний и величайший рубеж, наиболее сложная вещь из тех, которые мы пока обнаружили во Вселенной».

Но искусственный интеллект начался все-таки с математики. Думать о создании думающих систем, строить соответствующие теории начали в конце 30-ых – начале 40-ых годов прошлого столетия Алан Тьюринг, Клод Шеннон, Норберт Винер, Джон фон Нейман,

Фрэнк Розенблатт и др. Возникла необходимость решения таких задач, как расшифровка сообщений, отслеживание движущихся целей, быстрые расчеты тех или иных событий.

Вообще, сам термин «искусственный интеллект» был предложен 60 лет назад Джоном Маккарти. У нас подобные исследования получили признание в 60-ые – 70-ые годы. Основным центром этих исследований был механико-математический факультет МГУ имени М. В. Ломоносова. Именно тогда сложились ключевые идеи этого направления в науке. Среди них: формальные языки, распознавание образов, организация памяти, принятие решений, обучение, целесообразное поведение, оптимальная структура организации систем, сложность систем и т.п.

Особую роль сыграли участники семинара по кибернетике, который был открыт на мехмате в начале 50-ых годов. Я застал этот семинар и посещал его. Этот семинар был организован Александром Михайловичем Ляпуновым и Сергеем Всеволодовичем Яблонским. Я бы добавил сюда Олега Борисовича Лупанова, который долгое время был деканом мехмата. В работе семинара принимали участие Аксель Иванович Берг, Леонид Викторович Крушинский, Петр Сергеевич Новиков, Андрей Андреевич Ершов, Андрей Андреевич Марков–младший и др. Позже на факультете были открыты семинары по теории управляющих систем, сложности систем и алгоритмов, автоматов и роботов, распознаванию образов, базам данных и знаний, принятию решений. Возможно, импульс развитию данных направлений придали работы наших математиков, связанные с освоением космоса.

Практически сразу после появления компьютеров стали создаваться программы для решения центральных задач из самых разных областей. Они имитировали действия эксперта соответствующей области и получили название «экспертные системы». Таких систем очень много, некоторые из них могут успешно конкурировать с человеком, как, например, компьютерные шахматы. Столь высоких результатов подобные системы достигают не благодаря проникновению в логику человеческих действий, а, так сказать, методом грубой силы, за счет огромной производительности компьютеров для перебора вариантов.

Многие математики, специалисты в кибернетике и исследованиях искусственного интеллекта и психологии традиционно рассматривали создание программы для игры в шахматы как частный

случай для более общей проблемы когнитивных наук. Бурное развитие техники в конце 1990-ых годов сделало возможным построение программ, которые выбирают наилучший ход за счет перебора огромного числа вариантов. Обычные современные компьютеры анализируют сотни тысяч позиций в секунду и давно превосходят по уровню игру чемпионов мира. Суперкомпьютеры позволяют решать задачи, которые ранее казались невозможными. Например, на суперкомпьютере «Ломоносов» в 2012 году впервые в мире были просчитаны все варианты для шахматных позиций, содержащих семь фигур.

Однако принципы выбора хода, реализованные в современных компьютерных программах, существенно отличаются от тех, которыми руководствуется человек. По мнению первого советского чемпиона по шахматам Михаила Моисеевича Ботвинника, профессиональный шахматист, в отличие от программ переборного типа, обдумывая очередной ход, просчитывает только несколько десятков позиций. Это различие между решением задачи человеком и компьютером позволяет говорить о том, что задача моделирования мыслительных процессов шахматиста до сих пор не решена.

Попытки разработать алгоритм игры в шахматы, который отражает особенности логического мышления человека, а не добивается победы за счет вычислительного превосходства компьютера, предпринимались с начала 60-ых годов. Одним из таких подходов руководил Ботвинник, который также думал о проблеме компьютерных шахмат. Он работал над этим направлением вместе с нашими университетскими математиками, с моими учениками в том числе. В обсуждении он старался избегать самого слова «перебор», так как думал именно об интеллектуальном подходе. Построение модели мышления шахматиста является сложной задачей, которая, как и всякая задача в области когнитивной науки, объединяет несколько областей: психологию, математику, кибернетику, программирование.

На мехмате есть интересные результаты построения интеллектуальных компьютерных систем. Их называют «решателями». Автор этих решателей – Александр Сергеевич Подколзин. Это мой ученик, вместе с ним мы написали задачник для олимпиад по математике для студентов. Задачи охватывают разные предметные области: элементы дискретной математики, алгебру, элементарную алгебру множеств, элементарную геометрию, аналитическую геометрию, математический анализ, дифференциальные уравнения, теорию

вероятностей, ряд разделов элементарной физики, химии и шахматную теорию.

Обучающий материал для такого решателя содержит примерно 11 тысяч задач, из которых были построены около 39 тысяч приемов их решений. Задачи формулируются в решателях на специальном логическом языке, делятся на четыре типа: задачи на доказательство, на преобразование, на описание и на исследование. Рабочий цикл решателей состоит из сканирования текстов, просмотра встречающихся в них понятий, слов, обращения для текущих понятий к программе, объединяющей все закрепленные за понятием приемы. Это похоже на некое внутреннее логическое зрение этого решателя, которое обеспечивает контроль за развитием событий. Как и человек в процессе решения, система принимает «мысленное» рассмотрение объектов задачи, связей между ними для выработки плана дальнейших действий.

Такая интеллектуальная система построена, успешно функционирует, она популярна в математической среде, и, что самое главное, она решает 90 % задач из задачника школьной математики. Система легко поступила на механико-математической факультет, окончила несколько курсов на «хорошо» и «отлично», за секунды справляется с предложенными задачами, ее возможности намного превосходят все известные системы в этой области.

Это большое продвижение на том сложном пути, целью которого является создание искусственного интеллекта. Я буду рассказывать и о других подходах, но этот – один из самых наглядных и блистательных. Хотя эта программа не является прямой имитацией человеческого мозга, но, тем не менее, проучиться три курса на мехмате – это уже показатель высокоразвитого интеллекта.

Важная функция человеческого мозга – это обучаемость. Мозг самообучается постоянно, это наше свойство. Оно же является камнем преткновения на пути создания искусственного интеллекта. Как научить систему самообучаться? Мы часто говорим студентам, что надо научиться учиться, и помогаем в этом, но научить учиться компьютер – гораздо более сложная задача.

Работа над искусственным интеллектом требует глубоких знаний в области естественного интеллекта. Здесь крайне важна роль когнитивной психологии. В переводе с латыни «интеллект» – это «ум», «рассудок», «разум». Чаще всего интеллект отождествляют с мыслительной способностью человека. Однако человеческое мышление

имеет множество проявлений, и интеллект как способность не представляет собой единого целого.

Современная наука в формате логического мышления выделяет и другие типы интеллекта: биологический – это врожденные задатки к обработке информации, социальный – способности к различным действиям в социальной среде, невербальный – особенности восприятия, внимания и памяти, эмоциональный – способность считывать и понимать эмоции, намерения партнеров по общению, и многие другие типы. Но главное, что объединяет все эти типы интеллектов, – это способность человека эффективно адаптироваться к различным жизненным ситуациям, умение принимать правильные решения и быстро в непредсказуемых условиях решать задачи, которые требуют понимания неопределенности и риска.

Отсюда выводятся две большие задачи: уточнение феномена интеллекта и изучение механизмов, лежащих в основе разумного поведения. Это задачи многофакторные, они требуют учета большого количества переменных: общебиологических, нейрофизиологических, психологических и социальных наук. Кроме того, учитывая нейроадаптированный характер интеллекта, желательно проводить эксперимент в условиях реального времени.

В связи с этим я хочу упомянуть один проект. В рамках программы развития Московского Университета совместными усилиями различных ученых и факультетов создан многоцелевой программно-аппаратный комплекс, который включает в себя три экспериментальных блока: блок виртуальной реальности, блок регистрации движения, блок многоканальной регистрации активности мозга и периферической нервной системы. С помощью специально выделенной линии глобальной сети, комплекс, который находится на психологическом факультете, может подключаться к супервычислителю «Ломоносов», который находится на Воробьевых горах, т.е. от Моховой до Воробьевых гор.

Нашими учеными разработаны методологические основы для построения многофакторных психофизиологических моделей, которые описывают психофункциональное состояние человека: спокойствие, эмоции, стресс; это происходит в условиях виртуальной реальности, но максимально приближенной к реальной деятельности. Создан математический алгоритм, программное обеспечение для интегральной психофизиологической оценки изменений состояний

человека в режиме реального времени. Это еще один шаг в познании человеческого мозга.

Значительная часть предмета в поле искусственного интеллекта – различные приложения в области нейронаук, называемые нейроинтерфейсными технологиями. Одно из приоритетных направлений национальной технологической инициативы, выдвинутой нашим президентом В. В. Путиным, – нейронет, где этим технологиям уделяется значительное внимание. Интерфейсная технология основана на анализе электромагнитных сигналов мозга, мышц и движений глаз с целью расшифровки мысленных усилий человека относительно того или иного действия. Мысленные усилия человека должны быть расшифрованы и приведены в соответствующее движение с помощью мышц.

Это направление неслучайно получило наиболее эффективное развитие в Московском Университете, так как является междисциплинарным и требует взаимодействия ученых различных специальностей. Это кооперация нейробиологов, программистов, математиков и, в последнее время, экономистов и лингвистов. Нейроинтерфейсная технология открывает совершенно новый аспект дополненной реальности, когда к привычным мышечным и голосовым средствам коммуникации добавляются уже прямые каналы связи мозга с внешними исполнительными устройствами. Входным сигналом интерфейса мозг-компьютер является электрический сигнал мозга, но центральным процессом становится расшифровка самих мысленных усилий человека. Требуется расшифровать, что человек хочет, какой именно сигнал он послал. Если говорить о здоровом человеке, то нейроинтерфейсные технологии позволяют ему освоить мир, в котором отношения человека-оператора уже не будут ограничены двумя руками и голосом, а существенно пополняются возможностями прямого, от мозга, управления внешними устройствами.

Значительным потенциалом нейроинтерфейсные технологии обладают для здравоохранения, например, для реабилитации послеинсультных пациентов. На биологическом факультете разработаны и проходят клинические испытания первые образцы таких технологий для медицины. Это нейрокоммутатор, который обеспечивает контакт с внешним миром для пациентов, которые после травмы потеряли речь. Нейротренажеры через интерфейс типа мозг-компьютер управляются намерениями самих пациентов.

В лаборатории нейроинтерфейсов на моделях и на практике проходит испытания кресло-вертикализатор, которое позволяет лишенному подвижности пациенту в удобном ему режиме принимать вертикальное положение, что предотвращает обычные в таких ситуациях опасные застойные явления в организме. Кресло уже готовится к эксплуатации в Московском реабилитационном центре при МГУ.

В ближайших планах ученых МГУ в области нейроинтерфейсных технологий – построение математических моделей технологии нейрокоммуникации мозг-компьютер с разработкой специального мозг-машинного языка.

Я надевал на себя шлем нейроинтерфейса, конечно, он требует подготовки и тренировки, но выдающийся ученый биологического факультета Александр Яковлевич Каплан создал школу, благодаря которой мы являемся лидерами в мире по изучению этой технологии в рамках проекта нейроинтерфейса.

Еще один результат достигнут в области моделирования тактильного восприятия. Здесь я являюсь руководителем и автором этого проекта и прибора. Создан эндохирургический комплекс, способный воспринимать, преобразовывать, распознавать, анализировать, запоминать, передавать и воспроизводить тактильную информацию. Этот прибор не имеет мировых аналогов, востребован в медицине и уже производится. Речь идет о том, что среди человеческих чувств есть тактильное восприятие, которое редко изучается. Например, заходя в магазин за хлебом, мы так или иначе попробуем на ощупь, мягкий хлеб или нет. Это означает, что наш тактильный рецептор работает, когда мы убеждаемся, что хлеб мягкий.

Возникла группа, которой руковожу я, ее целью является создание медицинского прибора, позволяющего хирургу без сильного травмирования человеческих тканей проникать к больному органу, касаться его датчиком на конце введенного через прокол провода, который, в свою очередь, передает 3D-образ больного органа на экран. Вместе с изображением передаются данные о плотности и другие характеристики. Кроме этого, на столе у хирурга имеется манипулятор в виде груши, который тактильно прощупывается врачом с целью получения тактильной информации о воспаленном органе. То есть воспроизводится тактильный опыт, как если бы орган ощущался при полном вскрытии. Хирург видит трехмерное изображение, не наносит сильных повреждений пациенту и получает доступ

к частям внутренних органов, недоступных руке человека даже при полномерном разрезе.

Этот прибор производится на заводе «Сплав» в г. Туле. Недавно стало известно, что ведущая немецкая фирма по производству и продаже медицинской техники изъявила желание за огромную сумму купить как прибор, так и юридическое лицо, занимающееся производством и развитием этой технологии.

Есть еще одна тема, которую я хотел представить, и которой я лично занимаюсь с аспирантских лет. Это – тестирование качества управления движением в экстремальных условиях на динамических стендах виртуальной реальности. Ее цель – улучшение качества персонального управления движением в экстремальных условиях, например, управление в окрестностях орбитальной станции в невесомости. Здесь в условиях вестибулярных рефлексов есть рассогласование глаз от двигательного аппарата и вестибулярной системы человека, что существенно затрудняет управление аппаратом в космосе.

Этой лабораторией разработано программное обеспечение для согласования динамической и визуальной имитации. На аэрокосмическом салоне МАКС мы представили прототипы таких тренажеров. Когда я был аспирантом, ко мне обратился Георгий Тимофеевич Береговой, который тогда был руководителем центра подготовки космонавтов, с просьбой создания тренажера, который имитирует невесомость на земле.

В результате был выбран тренажер: мощная центрифуга, которая вращается с консолью 30 метров, она создает большие перегрузки вплоть до 30g. Далее, с помощью математических расчетов центрифуга запускается таким образом, чтобы вестибулярная система человека среагировала подобно тому, как она реагирует в космосе, то есть, чтобы всплытие кристаллов в вестибулярной системе было аналогично тому, как они всплывают в невесомости. Человек теряет земную ориентацию, но с помощью специальных скафандров, которые силой сжатия нижних конечностей направляют кровоток к голове, можно сделать так, чтобы человек не терял сознания, и кровь циркулировала аналогично тому, как это происходит в космосе.

Этот тренажер, созданный в 1980-ых годах, стал основным для тренировки космонавтов, отечественных и зарубежных. Мы же стали удостоены высшей премии, Государственной премии СССР за эту разработку, и считали, что этот успех связан с нейронаукой, поскольку

нами изучалось, как на земле можно заставить человека чувствовать невесомость, и таким образом вмешались в работу мозга.

Важно сказать, что в космосе человеческое зрение запаздывает на 1,5 секунды. Например, при повороте головы, результат этого действия человек видит лишь через полторы секунды. Это существенный промежуток времени, как показала авария с нашим космонавтом, когда он задел антенну. В ходе расследования изначально официальной причиной считалась недостаточная тренировка космонавта, но, как выяснилось некоторое время спустя, причиной допущенной ошибки была как раз задержка в получении зрительного сигнала мозгом на 1,5 секунды. За исследование этого феномена наша группа так же была удостоена государственных премий.

Сейчас наша группа занимается подготовкой прибора на станцию, которую запустят в 2017 году. Командиром этого экипажа будет Сергей Николаевич Рязанский, выпускник биологического факультета нашего университета. Сейчас мы занимаемся установкой прибора на саму станцию, который будет отслеживать конфликты зрительной и вестибулярной систем. Отчасти это также будет вкладом и в исследование искусственного интеллекта.

Хотелось бы отметить, что сегодня невозможно работать над исследованиями искусственного интеллекта без мощных супервычислителей. Они являются главными инструментами в этих исследованиях. Суперкомпьютеры – это огромные инженерные сооружения, работающие круглосуточно и требующие постоянного внимания. Например, суперкомпьютер «Ломоносов», расположенный в Ломоносовском корпусе МГУ, занимает два этажа целого крыла здания, площадь этого компьютера 1000 квадратных метров, вес 160 тонн, используется 50000 литров охлаждающей жидкости и 80 километров кабелей. Подобные системы исключительно мощные, они позволяют обслуживать тысячи пользователей одновременно, и эти пользователи географически могут быть расположены по всему миру.

Наша система создавалась в четыре этапа, начиная с 2009 года, таким образом, ее возраст уже 7 лет, в то время как средний возраст службы подобного оборудования 6 лет. Но наши ученые, учитывая сложность ситуации, умеют работать и сохранять жизнедеятельность и эффективность этого огромного комплекса за предельными сроками эксплуатации. Несмотря на такие условия, эффективность работы наших ученых, занятых на этом проекте, доказывают более 700 публикаций в ведущих мировых научных изданиях.

Сейчас мы имеем свободное пространство для установки дополнительного оборудования, чтобы достичь мощности суперкомпьютера в десятки петафлопс. Мы также выдвинули инициативу на достижение мощности, равной экзафлопсу, тем самым намереваясь лидировать в мире. Но даже сейчас по всем рейтингам мы находимся на достаточно высоком месте. И нет ни одного университета в мире, способного содержать подобную систему! В остальном мире подобные системы имеют только национальные центры США, Германии, Японии и Китая. Мощность нашей системы 10 мегаватт, сопоставима с потреблением электричества городом, а площадь с внешними охлаждающими блоками 5000 квадратных метров сопоставима с площадью завода.

Упомяну также наш спутник «Ломоносов», который сейчас находится на космодроме «Восточный». Запуск должен состояться совсем скоро, в апреле (27 апреля 2016 г. спутник «Ломоносов» был выведен на орбиту – прим. ред.). Этот спутник представляет собой научную космическую станцию, где, помимо всего прочего, установлен и прибор для изучения искусственного интеллекта. Кстати, запустить космическую станцию весом в тонну не берется ни один университет в мире.

Недавно при Московском Университете был создан Центр нейро- и когнитивных наук и уже приступил к работе. Создание Центра велось в рамках Института человека, который успел себя зарекомендовать как весьма функциональная площадка для междисциплинарных исследований.

Благодарю всех за возможность выступить на этой конференции.

ТРАНСФОРМАЦИЯ КУЛЬТУР: ОТ КЛАССИЧЕСКОЙ К ЭЛЕКТРОННОЙ TRANSFORMATION OF CULTURES: FROM CLASSICAL TO ELECTRONIC

米落诺扶 «文化转变：从古典时代到电子时代»

**Миронов
Владимир Васильевич**

*доктор философских наук, заслуженный профессор
МГУ, член-корреспондент РАН, декан философского
факультета Московского государственного
университета им. М. В. Ломоносова, г. Москва*

Mironov Vladimir V.

*Dr. Sci. (Philosophy), Honored Professor of Moscow State
University, Corresponding Member of the Russian Academy
of Sciences, Dean of the Faculty of Philosophy, Lomonosov
Moscow State University, Moscow*



Аннотация. Электронная культура и семиосфера. Электронная культура и общечеловеческая культура. Электронная культура: старые и новые формы диалога между локальными культурами. Трансформация культур. Междисциплинарные исследования ИИ – фактор интенсивного развития электронной культуры.

Ключевые слова: электронная культура, семиосфера, искусственный интеллект, диалог, трансформация культур.

Annotation. Electronic culture and the semiosphere. Electronic Culture and Human Culture. Electronic culture: old and new forms of dialogue between local cultures. Transformation of cultures. Interdisciplinary research of AI is a factor of intensive development of electronic culture.

Keywords: Electronic culture, the semiosphere, artificial intelligence, dialogue, the transformation of cultures.

Уважаемые коллеги! Виктор Антонович только что говорил о взаимосвязи факультетов университета, которые как исследовательские структуры занимаются одним делом – наукой, исследуя те или иные её аспекты. Одновременно он обратил внимание на интегративные процессы, происходящие в современной науке, когда учёные начинают работать на стыке разных предметных областей. В частности, и наша конференция выступает как пример такого типа, когда в обсуждении проблем участвуют нейрофизиологи, философы, математики и представители других научных направлений. В этом смысле, философия не является чисто гуманитарным знанием, как многие считают, а выступает в своей изначальной заданности как особого типа метанаука, которая находится на стыке разных наук, как гуманитарных, так и естественных, выполняя функцию, в том числе исследования философских оснований той или иной науки. Поэтому, мы могли бы, прослушав доклад ректора, лишний раз убедиться, что он здесь выступает не только как исследователь-математик, но по сути, как философ. К этому можно добавить целую серию статей, которые непосредственно касались философии, например, по поводу роли метафоры в научном знании или понимания философии как мудрости (доклад который он сделал на Российском философском конгрессе). Наконец, только что в прямом соавторстве с нашими коллегами во втором номере «Вопросов философии» вышла статья, посвященная социально-гуманитарной экспертизе.

Мой доклад носит немного иной характер. Искусственный интеллект – это явление все-таки особенное, связанное не только с наукой, но и с культурными аспектами, поэтому в своем докладе я попытаюсь рассмотреть социокультурные проблемы, связанные с развитием технологий. Неслучайно один из тезисов, с которого я начну, был представлен известным специалистом Антонио Паскуале, экспертом ЮНЕСКО, и звучит он так: «Нужно бояться не только спящего разума, порождающего чудовищ, но и чудовищности разума, страдающего от бессонницы».

Это проблема, которая действительно существует. Современные технологии и связанные с ними исследования совершают открытия настолько быстро, что ученые и общество часто не успевают к ним адаптироваться. Это снова заставляет нас задуматься о путях

Статья представляет собой расшифровку стенограммы доклада.

Исследование выполнено при финансовой поддержке РНФ: грант № 17-18-01620 «Антропологические трансформации в условиях сетевых медиа: новые режимы власти, знания, идентичности и коммуникации».

развития культуры, оценить ее нынешний этап, который можно назвать путем становления культуры электронной.

Для того чтобы двигаться в этом направлении, мы не можем говорить о культуре в целом, мы должны рассмотреть культуру как семиотическую систему, то есть как систему закодированных знаков, значений и смыслов. В этом случае становится понятно, что для носителей одной культуры смыслы другой культуры всегда закодированы мышлением, языком и часто могут быть непонятными для представителей другой культуры, с чем и связаны все парадоксы столкновения культур.

Культура выступает как локально замкнутая система, и ее познание, по сути, является проникновением в текст, потому что язык как система кодирования смыслов создает текст, который формирует память культуры. По блестящему выражению Юрия Михайловича Лотмана, язык представляет собой «память текстов». Главным инструментом кодирования выступает язык, который включает в себя историю создания и функционирования. Таким образом, язык – это код плюс его история. Если бы это был искусственный язык, то проблем бы не возникало. Наличие живой, обширной истории отличает живой язык от искусственного. Михаил Михайлович Бахтин заметил, что мы можем перевести слово любого языка на другой, но мы не можем перевести даже одного предложения. Это метафора, которая говорит о непереводаемости культурного смысла. К примеру, нельзя рассказать американцу анекдот о Штирлице, не вдаваясь в подробные объяснения.

Общечеловеческая культура представляет собой не культуру как таковую, цельную и единую, а является очень сложной системой взаимодействующих между собой локальных культур. Это сложный тип единства, единства разнообразного, а не тотального. Культура развивается за счет этого разнообразия. Индивид же включен в локальную культуру при рождении. Поэтому изменения, которые происходят в самой культуре, за время его жизни для него самого не очень заметны. То есть человек рождался, жил и умирал в одной и той же системе ценностей, которая модифицировалась незначительно, поскольку она не выходила за рамки его индивидуальной жизни. Ядро этой системы ценностей, некой конкретной культуры, на протяжении столетий тоже было неизменным. Это создает впечатление неизменности, вечности той системы, в которую человек погружен.

Диалог между локальными культурами реализовывался особым образом. Опять же, Ю. М. Лотман обозначал это термином «семиосфера», по аналогии с понятием Владимира Ивановича Вернадского «биосфера». В семиосфере роль живого элемента выполняет язык, а если точнее, то множество языков с их различающимися смыслами и формами культурного функционирования. В результате культуры пересекаются между собой как языковые множества. Область пересечения может быть большой, даже тождественной, но чаще всего она незначительна. Область же непересечения достаточно большая. Область тождества выполняет функцию проникновения, а область нетождественного представляется для нас, таким образом, центральным объектом исследования.

Академик Дмитрий Сергеевич Лихачев отмечал этот факт таким образом: диалог культур связан не только с познанием чужой культуры, но и с пониманием своей собственной. Ценность диалога связана не столько с пересекающейся частью, сколько с передачей информации между нетождественными частями. Это ставит перед нами серьезные проблемы.

С одной стороны, мы заинтересованы именно в общении в той ситуации, которая затрудняет общение, а в пределе может сделать его невозможным, поскольку возникают коллизии, когда культура блокирует информацию, которая для этой культуры нежелательна. Незнание, то есть область непересекаемого, требует адаптации, при которой желание и необходимость понимания увеличивает область пересечения, но процесс напряжения между культурами неизбежен. Этот процесс связан с тем, что мы реализуем особенности своей культуры, которые иногда не позволяют перенимать феномены другой и это является защитным механизмом сохранения своих культурных ценностей.

Становление же электронной культуры совпадает с процессами глобализации, которые приводят к появлению транснационального рынка, необходимости интеграции различных государств в различных транснациональных системах. Это осуществляется в условиях глобальной системы коммуникации, которая влияет на изменение диалогов между культурами. Деформируется характер диалога как семиосферы, и эта деформация оказывает влияние на компоненты самих культур, когда что-то не удается блокировать, или какая-то часть информации принимает абсолютизированное значение.

Объективность процесса глобализации не означает тотальную приемлемость этого явления. Хотя объективностью глобализации зачастую прикрываются негативные следствия: если с одной стороны глобализации выступают интеграционные процессы, то с обратной стороны будут процессы дезинтеграции. Термин «национальная дезинтеграция» ввел Александр Сергеевич Панарин, который задолго до современных проблем предсказывал появление этой ситуации.

Глобальная интеграция может реализоваться в системном тоталитаризме нового типа, в котором присутствуют совершенно уникальные возможности манипуляции сознанием человека с помощью современных технологий при внешней видимости демократического устройства. Этот образ глобальной империи, где внешне все в социальных сетях говорят что угодно, но на самом деле находятся в жесткой системе, где возможно детальное манипулирование, с помощью которого достигаются желаемые результаты. Поэтому ряд исследователей действительно называют это глобальной империей.

Культуры погружаются в глобальное коммуникативное пространство, которое меняет характер диалога между ними. Это можно назвать трансформацией культуры. Трансформация здесь понимается не просто как эволюционное изменение элементов системы, а как некий внутренний процесс, который отличен от революции и в большой степени сокрыт от наблюдателя, ибо реализуется за счет встраивания в нее чужеродных элементов, которые внешне не разрушают систему, а заставляют работать ее другим образом.

Здесь уместно привести аналогию с трансфекцией клетки. Это процесс, когда в клетку встраивается элемент чужеродной ДНК, клетка инфицируется, в результате чего запускается процесс трансформации, что приводит к изменению генотипа на генетическом уровне. В ряде случаев это позволяет обеспечить выживание организма, фактически изменив его изнутри, когда клетка сама начинает бороться с болезнями. Но в ряде случаев процесс трансфекции трудно контролировать, что приводит к мутациям. Важно отметить, что чем сложнее система, тем менее предсказуемым становится этот процесс.

Трансформация культур – это схожий, по аналогии, процесс привнесения в нее чужеродных элементов, запускающий механизм модификации культуры изнутри. Зачастую мы даже не осознаем, как эти процессы происходят. Если сегодня задавать вопрос детям «Кто такой Микки Маус?», то почти каждый ответит, но не каждый назовет героев русских сказок.

Модификация уже происходит и происходит незаметно. Условием этого как раз выступает глобальное коммуникационное пространство. Культура «атакуется» медиавирусами, как их называет Дуглас Рашкофф, американский медиавед, как и в случае с живыми организмами, в тех местах, где ослаблен иммунитет системы. Культура получает своеобразное «инфекционное заражение» через насаждение в нее стереотипов, которые не вытекают из ее истории и особенностей функционирования. Впитывания такого рода «инфекций» способны модифицировать человеческую культуру.

Сегодня модно говорить о толерантности, но при всех ее трактовках на биологическом уровне, например, толерантность будет означать принятие чужеродных элементов, что в случае неудачного перестраивания живого организма может привести к смерти. В рамках культуры толерантность будет означать терпимость к другому, и это очень хорошее качество, но это есть не что иное, как социально-психологическое состояние, которое не может длиться бесконечно. Более того, это сложное и некомфортное состояние как для того, кто терпит, так и для того, кого терпят. Это очень любопытно отражается в этимологии слова. Если мы посмотрим на синонимы «терпимости», то увидим: ожидать, переносить, цепенеть и т.д. Поэтому не может быть толерантности в отношении всего и вся. Это временные состояния, результатом которых должно стать включение другого в культуру в качестве равноправного объекта.

Удивительно, что Гёте использовал термин «толерантность» в своих рассуждениях таким образом: «Толерантность должна быть всего лишь временным убеждением, за ней должно следовать признание, поскольку терпеть значит оскорблять». Гёте пишет об этом задолго до религиозных и политических дискуссий. Толерантность нельзя навязать, ибо в таком случае реакцией станет безразличие к тому, что происходит, будь то на уровне индивида или на уровне общества. Абсолютизированная толерантность может укрепить чувство вседозволенности и забвения максимы «свобода одного кончается там, где начинается свобода другого».

Эта ситуация отражается в культуре в виде дихотомии «свой-чужой». Внутрикультурное здесь рассматривается всегда как имеющее большую ценность, чем то, что в нее проникает. Чужое может восприниматься как чуждое и в некоторых ситуациях как враждебное. Поэтому культура вырабатывает иммунитет к чужому, что является гарантией ее сохранения. Без этого иммунитета мы можем потерять

очень многое, и уже теряем. Внутри глобального коммуникационного пространства культуры все погружаются в единый информационный «мешок». Если раньше это была семиосфера, межкультурный диалог, то сегодня это информационный мешок, или инфосфера, которую пронизывают общие законы коммуникации, навязывающие нам некоторые стереотипы и решения, в том числе, политические.

Коммуникационные средства диалога превращаются в его сущность, которая очень хорошо выражается принципом «неважно, о чем говорить, главное говорить» (это у молодежи иногда называется «тусовкой»), коммуникация ради коммуникации без насыщения ее смыслом. Вместо диалога возникает некий глобальный монолог. Механизмом является использование всемирного языка, в качестве которого выступает английский язык. Глобальной культурой выбирается язык, который все равно относится к определенному государству, которое, в свою очередь, может проводить через него свои собственные культурные стереотипы. В связи с поверхностным, в основном, знанием чужого языка, общение в глобальной сети также носит поверхностный характер. К примеру, я сомневаюсь, что много людей посещает сайты, где представлены собрания средневековых библиотек на английском языке.

Далее происходит появление других культурных дихотомий. Например, дихотомия «откровенность/сокровенность», которая доминировала в классической культуре. Некоторые формы имели сокрытый характер, в наше же время они имеют форму открытого, доступного. Происходит, например, эротизация культуры, в которой содержится свобода выбора индивидуального удовольствия, выводя, таким образом, за пределы культурные «интимности».

Наносится удар по принципу завершенности в восприятии культурных единиц: завершенная симфония, завершенное архитектурное произведение, завершенный текст романа и т.д. Постепенно начинает господствовать клиповое сознание, основанное на поверхностном восприятии фрагментов реальности. Это сопряжено с потерей выявления сущностного, истинного характера явлений, когда теряется глубина исследований, и проблема истины как сущностной характеристики явления уступает свободной интерпретации, своеобразной игры, когда истиной можно считать все, что угодно.

Несмотря на увеличивающуюся продолжительность жизни, люди все равно не успевают адаптироваться к изменениям, которые происходят в культуре. Уменьшается пропорция между высокой и низовой

культурой. Низовая культура становится массовой. Например, в средние века, как пишет Бахтин, низовая культура господствовала во время карнавала, но не постоянно, как в наши дни. То есть сегодня карнавал становится постоянным действием, которое позволяют осуществлять современные технологии. Это приводит к тому, что главным становится не качество продукта творчества, а систематичность и способы его распространения. В этих условиях возникает феномен поп-культуры – доминирование низовой культуры в глобальном масштабе. Это хорошая среда для распространения медиавирусов, когда человек во всем ищет зрелищности.

Соответственно, меняются и характеристики нашего мышления. В электронной культуре изменяется роль экспертизы и роль эксперта. Электронная культура предоставляет бесконечные возможности поиска в почти безграничном поле информации. В связи с этим экспертиза часто сводится к поиску необходимой информации. Это приводит к изменению эпистемологического характера знаний, так как мы уже затрудняемся оценить объективность этого поиска информации. В качестве примера можно привести поисковые системы, которые подстраиваются под пользователя.

Свободный доступ к информации влияет на характер взаимоотношений между поколениями. В традиционной культуре всегда более взрослый человек выступал как обладатель большего количества информации. Сегодня же молодые люди имеют доступ к информации без посредника, и, в силу способности к более быстрому освоению технологий, доступ к информации для них открыт почти абсолютно. В свою очередь, это влияет на характер образовательного процесса. Если раньше преподаватель перед студентом был, прежде всего, носителем информации, то сегодня информацию можно получить без преподавателя. Возникают серьезные проблемы в рамках системы образования, благодаря открытому доступу к информации.

Электронная культура влияет и на само мышление. Уже сегодня ряд функций памяти человека оказывается ненужным. Например, молодое поколение не помнит номеров телефонов. Мир переходит к пользовательскому режиму, когда по любому запросу требуется всего лишь нажать кнопку, погуглить и т.д.

Это также меняет и характер принятия решений. Человек сначала осуществляет действия, а затем начинает задумываться. Определенное воздействие оказывается также на интуитивные способности человека, которые сокращают процесс научного поиска.

Исследователь не всегда может идти по пути бесконечного перебора вариантов решения той или иной задачи. Интуиция часто позволяет ученому выбрать правильный вариант решения без полного перебора. Ученые часто говорят, что закон Кеплера никогда не мог бы быть открытым с помощью ЭВМ, так как перебираемые факты могли бы закрыть своим множеством догадку о том, что земля вращается вокруг солнца по эллипсу.

Электронная культура порождает целый ряд коммуникативно-психологических проблем. Например, проблема адаптации человека к возрастающему количеству информации и пр. В каком-то смысле идет возврат к рукописному типу культуры, так как идет доминирование виртуальных образов, имеющих очень сильное воздействие.

Электронная культура делает жизнь человека более комфортной. Человек не может отказаться от благ, предоставляемых развитием науки и техники, даже если он, например, знает о тотальной слежке с помощью электронных устройств и Интернета: банковские карты, счета, социальные сети, покупки через Интернет и пр. Это создает благоприятную среду для манипуляций.

И, конечно, несколько слов о философии. Философия – это самосознание культуры, а значит, она отражает те процессы, что происходят в культуре. Гегель, к примеру, мог появиться только в немецкой культуре и именно в то время. Происходящие культурные процессы становятся объектом научного исследования. Сейчас появляются философские исследования конкретных фильмов и сериалов, как объектов культуры.

Грань между реальным философским содержанием и его имитацией исчезает. Постмодернизм, в частности, порождает имитирующие формы философии, где центральным стержнем выступает критическое отношение к рациональным ее формам, тексту и т.д. И в этих условиях философские размышления по поводу дальнейшего развития культуры как раз представляются важными как никогда. Задуматься о возможностях управления глобальными процессами в культуре и истории – одна из центральных задач в философии. Также важно предупреждение о тенденциях, которые связаны с научно-техническим прогрессом. Они несут не только положительные, но и отрицательные последствия для человека и общества.

ФИЛОСОФИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, КОГНИТИВНЫЕ ИССЛЕДОВАНИЯ

PHILOSOPHY, ARTIFICIAL INTELLIGENCE, COGNITIVE RESEARCH

列可多斯及 «哲学，人工智能，认知研究»

**Лекторский Владислав
Александрович**

*академик РАН, доктор философских наук, профессор,
заведующий сектором теории познания Института
философии РАН, г. Москва*

Lektorsky Vladislav A.

*Academician of the Russian Academy of Sciences, Doctor
of Philosophy, Professor, Head of the Knowledge Theory
Sector of the Institute of Philosophy of the Russian
Academy of Sciences, Moscow*



Аннотация. Рассматривается история формирования научного направления «искусственный интеллект» в контексте развития когнитивных исследований. Отмечаются этические проблемы, связанные с разработками в области искусственного интеллекта. К их числу относится проблема свободы воли у искусственной интеллектуальной системы и вопрос о негативном воздействии на своих создателей «сильного» искусственного интеллекта.

Ключевые слова: философия искусственного интеллекта, когнитивная наука, этические проблемы искусственного интеллекта.

Annotation. The history of the formation of the scientific direction «artificial intelligence» is considered in the context of the development of cognitive research. Ethical problems associated with developments in the field of artificial intelligence are noted. These include the problem of free will in the artificial intellectual system and the question of the negative impact on their creators of «strong» artificial intelligence.

Keywords: philosophy of artificial intelligence, cognitive science, ethical problems of artificial intelligence.

Тема моего доклада очень большая, я попробую сказать только о самых важных вещах. Прошло 60 лет с тех пор, как возникло направление исследований и разработок, получивших название «искусственный интеллект». С тех пор утекло много времени. До того, как возникло это направление, существовала кибернетика, сейчас о ней уже почти не говорят. У нас одно время кибернетику ругали, потом признали и даже превозносили. Под редакцией академика А. И. Берга издавалась серия книг «Кибернетику – на службу коммунизму», это были очень хорошие книги. Затем, уже из движений кибернетики, информатики возникло направление под названием «искусственный интеллект», а спустя еще 10 лет, в 60-е годы – когнитивная наука. Это взаимосвязанные вещи, но не совсем одно и то же.

Сначала речь шла о том, что можно имитировать некоторые мыслительные процессы человека и достигать тех же результатов, что и человек, но, может быть, другими средствами. И иногда это можно делать лучше, чем делает человек. Эта идея стала воплощаться в жизнь, и у нас также была подхвачена. Стали разворачиваться дискуссии на тему: «Могут ли машины мыслить?» Связаны они были с тем, что известный математик Алан Тьюринг опубликовал статью с таким названием («Может ли машина мыслить?»). В статье сформулировано положение, которое было названо потом тестом Тьюринга. Опубликовал он свою статью в журнале «Mind» – главном философском аналитическом журнале, потому что проблема это именно философская. Тьюринг предлагает такой тест. В одной комнате сидит человек, а в другой машина. Но человеку неизвестно, кто находится в закрытой для него комнате: то ли человек, то ли искусственный механизм. Человек задает в соседнюю комнату вопросы, и, если на все вопросы неизвестное существо отвечает правильно, осмысленно, и человек в результате долгой беседы не может понять, кто ему отвечает, машина или человек, значит, это нечеловеческое существо – машина – мыслит не хуже человека.

Об этом тесте потом много писали, дискуссии об этом до сих пор идут. Многие возражали против самой формулировки этого теста: не только философы, но и ученые. Например, из тех, кто не принимает тест Тьюринга и идею, что можно уподобить познавательные и психические процессы некоторым вычислительным процессам, известный физик и математик Пенроуз. Сразу же философские вопросы не могли не возникнуть. Ведь это вызов человеку. До сих пор казалось,

что если что-то и отличает человека от всех других живых существ, так это то, что он может мыслить.

Об этом философы писали всегда, начиная с античности. Человек – это разумное существо. Позже свобода воли еще была к этому присоединена, это уже Средние века и Новое время. Разум плюс свобода воли – то, что не может быть ни у кого, кроме человека. О том, что создать можно какую-то машину, которая будет имитировать некоторые действия человека, писал еще Декарт. Но эта машина не будет обладать разумом и свободой воли. Вызов был сформулирован самой идеей искусственного разума, идеей о том, что машина может что-то делать лучше человека. Сегодня мы знаем, например, что играть в шахматы машина может куда лучше человека. Идея «искусственного – естественного» сейчас стоит более остро, чем когда бы то ни было, острее, чем даже во времена Тьюринга. Сегодня выдвигается некоторыми теоретиками идея о том, что можно искусственно создать всё то, что возникло естественно, и даже то, что природа создать не смогла. Это уже вызов космосу. Некоторые теоретики считают, что человечество на новом этапе может превратиться в демиурга вселенной, управлять эволюцией.

Примерно спустя 10 лет после оформления направления под названием «искусственный интеллект» возникла когнитивная наука. Она была также создана в США. Если автором термина искусственный интеллект был Джон Маккарти, то понятие когнитивной науки ввел Марвин Минский, известный ученый из Бостона. Идея когнитивной науки состояла в том, что если искусственный интеллект ставит цель разрабатывать такие механизмы, с помощью которых машины могут делать то, что делает человек тем или иным способом, то, опираясь на эту идею, надо изучать процессы естественного человеческого мышления, познания, вообще всех психических процессов. Опираясь на компьютерную метафору, можно понимать все психические процессы как вычисление на некоторых ментальных репрезентациях. Такие ментальные репрезентации могут быть и у такого существа, у которого отсутствует мозг. Итак, в основе когнитивной науки лежат две идеи: во-первых, все психические или ментальные процессы можно понять как процесс некоторых вычислений, во-вторых, эти вычисления возможны тогда, когда есть некоторая структура, особая ментальная репрезентация. Если это так, то с точки зрения основателей когнитивной науки к этому сводятся все психические процессы: не только мышление, но и восприятие, боль, любовь. При таком

понимании ментальных процессов для их исследования мозг можно не изучать.

Философы с самого начала этого движения не только истолковывали уже полученные учёными данные, но и пытались методологически осмыслить новое движение. Один из первых был Хилари Патнэм, известный американский философ. Он сформулировал идею о том, что все психические процессы можно понять как некоторые функции, это идея функционализма. А функция может быть осуществлена на любом носителе. У человека это происходит на мозге, у инопланетного существа это может быть не мозг, а какая-то другая структура. У машины – в принципе то же самое. Мы говорим: «боль», и нам не важно, кто и как чувствует боль, а важно, что боль имеет функциональный смысл, сигнализирует, что что-то не в порядке. У другого существа сигнализация боли будет происходить иным способом, чем у человека, но функция этого ментального состояния будет той же самой. Поэтому, изучая эти функции в рамках когнитивной науки, мозг можно не изучать.

В когнитивную науку вошли как её составные части искусственный интеллект, когнитивная психология (психологи это подхватили сразу), когнитивная лингвистика (Ноам Хомский подошел к близким идеям). И так это дело шло в течение некоторого времени: психологи выделяли разные функциональные блоки переработки информации, но не пытались разрабатывать программы для машины, они считали, что в принципе это можно сделать, можно всё это представить как вычисления, а какие именно это вычисления – материал для будущих разработчиков в области искусственного интеллекта.

Эти идеи были очень популярны в 70-е годы, начале 80-х годов прошлого столетия, а потом оказалось, что всё происходит как-то не так и что всё упирается в понимание фундаментальных оснований когнитивных исследований, которые носят философский характер. Известный теоретик когнитивной науки, американский психолог и лингвист, философ Джерри Фодор однажды сказал, что многие идеи появляются в истории дважды: сначала в философии, а потом в когнитивной науке. Он написал статью «Методологический солипсизм как основная идея когнитивной науки».

Идея статьи была такой. Для того чтобы понять психические процессы, достаточно изучать то, что происходит внутри устройства, перерабатывающего информацию. Для человека это переработка информации внутри мозга. На человека, конечно, идет воздействие

извне, он живет в мире, но, получив информацию из мира, мозг начинает её обрабатывать по своим программам, алгоритмам, правилам, всеми способами, которые у него есть. И потом какой-то выход выдает, который влияет на поведение. Для того чтобы понять, как работает мозг, не нужно изучать, как устроен внешний мир и как он воздействует на человека. Более того, при переработке информации происходит работа лишь с формальными синтаксическими свойствами информации по правилам, которые записаны на языке мысли. Д. Фодор предложил идею об особом врождённом языке мысли (эта идея отсылает к Декарту). А какое содержание информации перерабатывается, можно и не знать для того, чтобы понять, как работает мозг. Солипсизм понимается в данном случае не в том смысле, что отрицается существование внешнего мира. Внешний мир есть, но с точки зрения Фодора его можно не принимать во внимание для того, чтобы понять, что такое ментальные феномены и процессы.

Но первоначальная программа когнитивной науки столкнулась с большими противоречиями и трудностями.

Началась критика когнитивной науки в том виде, в котором она сначала была задумана. Сейчас то, как существует и развивается когнитивная наука – это не то, что было вначале. Во-первых, выяснилось, что без изучения мозга не обойдешься, потому что важно учитывать не просто способы переработки информации, но и природу носителя этой переработки. И, как выяснилось, не любой носитель допускает любые возможности для переработки любой информации. Во-вторых, выяснилось, что солипсистское понимание психики Дж. Фодора тоже невозможно, и пришлось эту установку менять, и обратиться к другой парадигме когнитивной науки. Последняя при своём возникновении отталкивалась от англо-американской философской аналитической традиции. А когда первоначальные представления когнитивной науки подверглись критике, пришлось обратиться к другой традиции – европейской, идущей от феноменологии Гуссерля и Мерло-Понти. Теперь большинством исследователей, работающих в этой области, признаётся, что нельзя ничего понять в сознании, психике, в ментальных процессах, если мы не учтем, что познающее существо включено в мир и находится в постоянном взаимодействии с ним. Даже восприятие – это не просто получение информации, а постоянное взаимодействие с миром, и событие внутри мира.

В этой связи я хотел бы еще раз обратиться к идее солипсизма. Д. Беркли солипсистом не был, но из его учения легко сделать

вывод, что существую только я и мои ощущения. Идея сумасшедшая, но попробуйте её опровергнуть. Известен исторический эпизод, когда Беркли гулял по улице со своим другом доктором Джонсоном, и последний, ударив по камню, сказал: «Смотри, я опроверг твой тезис о том, что существуют только мои ощущения: я толкнул камень, он в результате покатился, при этом я почувствовал (через осязание) силу сопротивления камня. Значит, камень существует как внешняя реальность, которая может воздействовать на мои ощущения и поэтому существует независимо от них». Позже Э. Кондильяк провёл идеальный эксперимент со статуей, которую наделяют разными ощущениями: обонятельными, зрительными, звуковыми. Оказывается, что в этих случаях для статуи невозможно отличить, что существует на самом деле, а что ей только кажется. Всегда в этих случаях иллюзия возможна: и обонятельная, и вкусовая, и зрительная, и слуховая. Но вот когда статую наделяют осязанием, тогда возникает ощущение сопротивления мира, и статуя получает возможность различить то, что находится внутри неё и что существует в реальном мире. Но ведь осязание предполагает активные действия познающего существа, использующего свои телесные возможности: это, в частности, конечности у животных, руки и ноги у человека.

Если мы принимаем в когнитивной науке т.н. «телесно-воплощённый» подход и тесно с ним связанный деятельностный подход (*embodied and enacted approaches*), то изменяется понимание целого ряда проблем. В частности, придётся признать, что мир будет восприниматься по-разному в зависимости от того, как устроено познающее существо. Один и тот же реально существующий мир будет выступать как разный для пчелы, для блохи, для волка и для человека.

В 1930-е годы, когда был на взлёте логический позитивизм, Рудольф Карнап написал известную статью «Критика метафизики с помощью логического анализа языка», в которой объектом критики был Хайдеггер, высказывания которого трактовались Карнапом как вершина бессмыслицы. Но сейчас некоторые специалисты в области когнитивной науки обращаются как раз к Хайдеггеру как источнику идей в их области. Я уже говорил о том, что ныне феноменология вступает в тесное взаимодействие с когнитивной наукой. Издаётся популярный журнал «Феноменология и когнитивная наука». Разработанные в нашей философии и психологии деятельностный подход и культурно-историческая теория психики и познания становятся востребованными современной когнитивной наукой.

Вот ещё одна проблема, горячо обсуждаемая в современных когнитивных исследованиях: свобода воли. Как мы помним, с точки зрения классической философской традиции именно разум и свободная воля отделяют человека от всех других живых существ. Между тем, известный специалист в области когнитивной нейронауки Бенджамин Либет провёл серию экспериментов, которые вроде бы показали, что никакой свободы воли в действительности у человека нет. По этому поводу идёт большая дискуссия среди учёных и философов. Я не буду останавливаться на этой дискуссии, скажу только, что с моей точки зрения эксперименты Либета вовсе не доказывают отсутствие свободы воли. Для меня сейчас важно другое: сегодня когнитивные исследования создают вызов многим традиционным представлениям, укорененным в классической философской традиции и в самой культуре.

То, что сейчас делается в когнитивных науках и связанных с ними нано-био-инфо-когнитивных технологиях, это, конечно, вызов. По сути дела это взрыв той культуры, в которой человек существовал несколько тысяч лет. С этими исследованиями и технологиями, в частности, связана популярная сегодня идея трансгуманизма, идея создания такого искусственного существа, которое жить будет лучше, чем современный человек, есть меньше, а может быть, и умирать никогда не будет. Это можно считать фантастикой, но некоторые серьёзные люди об этом серьёзно пишут, есть связанное с этой идеей целое движение в мире и в нашей стране. Как говорил в своё время Иван Владимирович Мичурин: не надо ждать милостей от природы, нужно взять их, создать то, что не может создавать природа сама по себе, эволюцию под контроль поставить. Но в этом же и идея трансгуманизма. Якобы человек должен уступить место некоему новому существу, постчеловеку, и тогда начнется новая ступень всемирной эволюции.

Я понимаю эти проблемы совсем иначе. Конечно, человек – это несовершенное существо, у него много недостатков, он может уступать машине во многих отношениях. Но человек – это человек, и, пытаясь что-то в нём улучшить, мы часто что-то ухудшаем, и при этом можем создать такое существо, которое будет чудовищем и породит новые опасности. Были правы Хайдеггер и другие, когда говорили о том, что, если существо бессмертно, то ему не нужны мужество, страдание, любовь. Этот постчеловек будет просто нелюдью. Об опасности этого, например, написал известный физик Хокинг. Надеюсь,

что постчеловека никогда не будет. Но вот, что на самом деле есть уже сегодня: мы незаметно вползаем в новую ситуацию, связанную с новейшими технологиями, которая уже сейчас влияет на человека, создаёт новые проблемы

Возникает новая среда, в которой мы сейчас живем и незаметно меняемся. С одной стороны, жить становится лучше. Например, Интернет появился, теперь можно как бы преодолевать пространство и время, получить доступ к самым разным информационным ресурсам. Но, с другой стороны, лёгкая возможность получать готовую информацию отрицательно влияет на культивирование творческих способностей. В советское время многие психологи писали о том, что школа должна учить мыслить, а сейчас психологи и педагоги криком кричат о том, что современным детям часто не нужно мыслить, Интернет их от этого отучает, ведь вся информация в нём уже есть. Правда, в принципе можно и Интернет использовать для развития творческого мышления. Появились разработки по обучению мыслить с помощью Интернета, но в любом случае обучать творческому мышлению сегодня стало сложнее. А вот другая проблема. Сегодня с помощью вживлённых в тело человека чипов можно контролировать его передвижения и даже поведение. Особенно это становится актуальным в период террористической опасности и может использоваться государственными структурами. Люди могут пойти на это, и даже наверняка пойдут. Но в этом случае не всегда будет легко провести грань между заботой о безопасности и контролем за поведением просто в интересах государственной машины. За безопасность можно заплатить ценой индивидуальной свободы. Это реальная проблема.

Вот сколько сложных вопросов возникло перед человеком с тех пор, как возникла идея искусственного интеллекта и начались разработки в этой области.

Сегодня человек стоит на краю пропасти. Он может оттолкнуться от этого края и высоко взлететь, стать лучше во всех отношениях. Такая возможность, бесспорно, есть. Но человек может и упасть в пропасть. Идея искусственного интеллекта судьбоносна. И при обсуждении современных проблем, возникших в связи с развитием информационных и когнитивных технологий, без философии никак не обойтись. Анализ субъективности, сознания, познания свободы воли, человеческих ценностей – то, чем всегда занималась философия, это сегодня передний край развития науки в целом и необходимое условие сохранения и развития человека.

СОЗНАНИЕ, МОЗГ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НОВЫЕ ПРОБЛЕМЫ

CONSCIOUSNESS, THE BRAIN, ARTIFICIAL INTELLIGENCE: NEW PROBLEMS

读布落扶其《意识，脑，人工智能，新问题》

Дубровский Давид Израилевич

*сопредседатель НСММИ РАН, доктор философских
наук, профессор, главный научный сотрудник
Института философии РАН, г. Москва*

Dubrovsky David D.

*Co-chairman of the RAS SCMAI, Doctor of Philosophy,
Professor, Chief Researcher, RAS Institute of Philosophy,
Moscow*



Аннотация. Исследование естественного интеллекта, психических процессов и феномена сознания – важнейшая задача современной науки и практики, источник новых подходов в разработке информационных и когнитивных технологий, а тем самым необходимое условие продуктивного развития всего комплекса НБИКС-конвергенции.

Ключевые слова: искусственный интеллект, естественный интеллект, сознание, нейронаука, когнитивные исследования, информационные технологии, НБИКС-конвергенция.

Annotation. The study of natural intelligence, mental processes and the phenomenon of consciousness is the most important task of modern science and practice, the source of new approaches in the development of information and cognitive technologies, and thus the necessary condition for the productive development of the entire complex of NBICS convergence.

Keywords: Artificial intelligence, natural intelligence, consciousness, neuroscience, cognitive research, information technology, NBICS-convergence.

Прежде чем приступить к докладу, здесь было бы уместным коротко рассказать о Научном совете РАН по методологии искусственного интеллекта, который совместно с философским факультетом МГУ является организатором этой конференции. Наш Совет был создан в марте 2005 года, и сегодня мы можем отметить ровно 11-летний срок его деятельности. Решение о создании Совета было принято на первой Всероссийской конференции «Философия искусственного интеллекта», которая состоялась на базе Московского государственного института электроники и математики (18-20 января 2005 г.). И можно отметить некоторую преемственность между той конференцией и нынешней.

За 11 лет Советом и его секциями проведено более 100 ежемесячных теоретических семинаров по междисциплинарным проблемам искусственного интеллекта, в которых принимали участие ведущие специалисты в области компьютерных дисциплин, информатики, логики, психологии, нейрофизиологии, лингвистики, математики, робототехники, философии. Советом издан значительный объем литературы по философским, методологическим, теоретическим проблемам искусственного интеллекта и его соотношениям с естественным интеллектом, проведено большое число научных конференций, симпозиумов и круглых столов. Советом создано 15 региональных отделений в ряде крупных научных центров страны (Санкт-Петербургское, Архангельское, Пермское, Вологодское, Ивановское, Новосибирское, Дальневосточное, Московское-областное и др.).

В последние годы структура и проблематика Совета расширены за счет создания восьми специализированных секций, работающих на базе крупных учебных и научных заведений Москвы. Две из них созданы и успешно работают в Московском государственном университете им. М. В. Ломоносова. Это секция «Нейрофилософия», которая провела уже 21 семинар (руководители проф. В. Г. Кузнецов и доц. А. Ю. Алексеев) и секция «Управление знаниями» (руководитель проф. Ю. Ю. Петрунин). Считаю долгом отметить главную роль в создании секций координатора научных программ Совета Андрея Юрьевича Алексеева и его большой вклад в организацию нашей конференции.

В составе Совета более 10 лет активно работает молодежная секция, которой руководство Совета уделяет особое внимание. Она также имеет свои отделения во многих городах страны. 9 лет подряд мы ежегодно проводим Всероссийскую научную конференцию

студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации», которая собирает участников из десятков городов страны. Издано 16 томов материалов этих конференций.

Надо подчеркнуть, что вся эта большая работа проводится на общественных началах, без какой-либо материальной поддержки Академии наук и государства. И это при том, что наш Совет единственная структура в стране, которая на протяжении 11 лет систематически обсуждает и разрабатывает междисциплинарную проблематику искусственного интеллекта, когнитивных наук и конвергентного развития НБИКС (нанотехнологий, биотехнологий, информационных, когнитивных, социогуманитарных технологий и соответствующих им областей научного знания).

Может быть теперь ФАНО, которое так много говорит о междисциплинарности и придает ей столь большое значение, обратит внимание на нашу деятельность и окажет нам поддержку?

Разрешите теперь перейти к докладу.

С точно таким докладом я выступал ровно 11 лет тому назад на первом заседании Совета. С того времени наука сделала большой шаг вперед, но эта тематика по-прежнему сохраняет высокую актуальность. Существенно возросла взаимозависимость между исследованиями мозга, сознания и разработками в области искусственного интеллекта.

Исследование естественного интеллекта, психических процессов и феномена сознания остается важнейшей задачей современной науки и практики, источником новых подходов в разработке информационных и когнитивных технологий, а тем самым необходимым условием продуктивного развития всего комплекса НБИКС-конвергенции. Постоянный интерес к взаимосвязям искусственного интеллекта с естественным интеллектом, к вопросам междисциплинарности в проблематике искусственного интеллекта характерен как раз для тех исследователей, которые нацелены на осмысление стратегических задач развития искусственного интеллекта, кто отдает себе отчет в том, что масштабные новаторские решения зависят от теоретических достижений, а последние в существенной степени иницируются, стимулируются основательным эпистемологическим и методологическим размышлением и анализом (о чем свидетельствует история науки).

В последние десятилетия нейронаука добилась существенных результатов в изучении психических процессов и феноменов сознания. Эти результаты были связаны с использованием методов визуализации нервных процессов, «картирования мозга», отведения сигналов от отдельных нейронов, методов анализа нейropsychических корреляций с применением нейрокомпьютерных интерфейсов и средств информатики.

На повестке дня, однако, совершенствование и использование новых методов, создаваемых на основе достижений нанотехнологий, оптохимии, биоинформатики, нейрорадиологии, генной инженерии, обладающих гораздо большей разрешающей способностью визуализации, открывают возможность фиксировать и изучать те процессы, которые одновременно совершаются в больших популяциях нейронов (насчитывающих десятки, сотни тысяч и миллионы нервных клеток). Именно в таких крупных популяциях нейронов, образующих многомерные сетевые динамические структуры, возникают «эмерджентные свойства» нейронных процессов, которые определяют психические состояния и поведенческие акты. При этом используются новейшие средства информатики, создающие возможность анализировать и систематизировать огромные потоки данных, вплоть до петабайтов информации.

Хотелось бы выделить два перспективных фундаментальных направления нейронауки. Это исследования «зеркальных» нейронов и систем в головном мозге (1) и исследования специфики мозговых информационных процессов, представляющих явления сознания (2), того, что именуют «Чтением мозга» («Brain-reading»). Более точно это направление может быть названо нейрокриптологией, ибо оно ставит задачу расшифровки мозговых нейродинамических кодов явлений субъективной реальности. Эта задача была поставлена вслед за расшифровкой генетического кода и открывает столь же масштабный, стратегически значимый этап в развитии науки и технологий преобразовании человека.

Важно отметить, что эти исследования проводятся, как правило, на основе информационного подхода, который создает возможность для построения моделей, объясняющих связи нейронных процессов и явлений субъективной реальности и их каузальных функций как формы информационной причинности. Понятия «информационного процесса», «паттерна информации», «кодовой нейроструктуры», «кодирования», «перекодирования», «расшифровки кода» органически

используются нейронаукой в целях описания и объяснения деятельности мозга.

Это особенно четко видно и в концепции зеркальных нейронов. Она позволила пересмотреть классические представления о связи сенсорных и моторных функций в деятельности головного мозга. По словам основателя этого направления Джакомо Ричцоллатти, зеркальный механизм одновременно осуществляет «кодирование сенсорной информации в моторных терминах» (1, с. 122), «фиксирует интенциональный аспект действия, общий для наблюдателя и исполнителя» (там же, с. 118), моментально перекодирует «наблюдаемые действия в двигательные коды» (там же, с. 129) и тем самым обеспечивает понимание действий другого. Зеркальные системы основательно объясняют понимание эмоций. Тем самым они формируют фундаментальный механизм коммуникации, социальных взаимодействий между индивидами.

Результаты исследований зеркальных систем имеют существенное значение для проблематики искусственного интеллекта. Несмотря на то, что основной массив разработок в этой области идет по пути, так сказать, «независимому» от исследований мозга (ставится интеллектуальная задача, и она решается путем изобретательного программирования, использования все более изощренных логических и математических методов), стратегические перспективы развития искусственного интеллекта связаны с углублением понимания тех нейродинамических самоорганизующихся структур головного мозга, которые обеспечивают функционирование естественного интеллекта.

В рамках проблематики искусственного интеллекта обычно ставится вопрос: возможно ли воспроизведение психических явлений искусственным путем на небиологическом субстрате? На мой взгляд, такая возможность теоретически мыслима. Основанием для этого служит принцип изофункционализма систем (А. Тьюринг), согласно которому одна и та же функция (система функций) может быть воспроизведена на разных субстратных началах, и принцип инвариантности информации по отношению к физическим свойствам своего носителя, согласно которому одна и та же информация может быть воплощена и передана разными по своим физическим свойствам носителями (т.е. кодироваться по-разному)

Теоретически мыслимо воспроизведение психических явлений на основе небиологических систем, достигающих такого уровня

структурно-функциональной динамической организации, которая аналогична мозговым информационным процессам, производящим качество субъективной реальности.

Эти теоретические вопросы серьезно обсуждаются в последние годы не только философами (3), но и крупными представителями нейронауки (4), особенно в связи с перспективами НБИКС-конвергенции, которые намечают качественно новые прорывы в развитии искусственного интеллекта и информационных технологий в целом.

Важные результаты достигнуты в последние годы в области «Чтения мозга», нейрокриптологии. Прежде всего это результаты расшифровки мозговых кодов зрительных образов (Мияваки, Камитани, Нишимото, Гэлбрейт и др.). Установлены довольно точные нейродинамические корреляты многих психических явлений.

Проблема расшифровки нейродинамических кодов психических явлений является особой задачей, так как включает герменевтический аспект – вопрос понимания информации. Здесь требуется специальный анализ ряда теоретических и методологических вопросов (см.: 5), должен использоваться опыт криптологии, лингвистики, этнографии. Кодирование и декодирование – ключевое звено информационных процессов в биологических, социальных и технических системах.

Расшифровка кодов представляет собой центральную по своей значимости задачу во многих науках – как в естествознании, так и в социогуманитарных дисциплинах. Ее решение позволяет объединить в едином концептуальном плане естественнонаучные и социогуманитарные подходы в исследовании человекообразных самоорганизующихся систем.

Я хотел бы поэтому подчеркнуть фундаментальный характер проблемы расшифровки кода не только в нейронауке (или в информатике, психологии, лингвистике, этнографии и т.д.), но как широкую биосоциальную, социокультурную, коммуникативную, технологическую проблему. Это можно проиллюстрировать на примере трех эпохальных периодов нашей цивилизации, начало которых было связано с решением задачи расшифровки кода.

1) Зарождение компьютерной эры связано с теоретическими разработками А. Тьюринга, в частности с его знаменитой «Машиной Тьюринга». К этому, конечно, были причастны многие выдающиеся ученые, внесшие весомый вклад в создание и совершенствование компьютера (прежде всего Дж. фон Нейман и др.) Но первые

практические шаги были сделаны А. Тьюрингом, когда он решал задачу расшифровки кода знаменитой «Энигмы» и особенно последовавшей затем расшифровки кода «Лоренц», который был намного сложнее «Энигмы». С помощью кода «Лоренц» шифровались и передавались сообщения высшего германского командования вермахта и приказы самого фюрера. В середине 1944 г. код «Лоренц» был также расшифрован А. Тьюрингом, и можно представить, какое преимущество это дало союзникам на последнем этапе войны. Но для решения этой задачи Тьюрингом была создана знаменитая вычислительная машина «Колосс» – первый прообраз компьютера.

Это положило начало компьютерной эре, развитию информационных технологий. Хорошо известно, какие глобальные преобразования нашей цивилизации на основе информационных технологий за этим последовали.

2) Второй период, не менее значимый по своим масштабам и последствиями для земной цивилизации, связан с расшифровкой генетического кода и генома человека (генная инженерия и биотехнологии, создание искусственной клетки Вентер и др.), возникли новые захватывающие перспективы и новые столь же масштабные риски и угрозы человечеству.

3) Сейчас мы вступаем в третий подобный период – расшифровки мозговых кодов психических явлений, сознания, субъективной реальности человека. Как уже отмечалось, это направление исследований развивается весьма успешно, достигнуты многие стойкие нейродинамические корреляты психических явлений, которые служат медицинским целям, развитию нейротехнологий. На этой основе созданы, например, управляемые мыслью протез руки, инвалидная коляска и даже экзоскелет для парализованного человека, с помощью которого он сам может ходить, многие другие устройства такого типа. Недавно английский нейрофизиолог Адриан Оуэн смог установить общение с больной, пребывающей пять месяцев в вегетативном состоянии. У таких пациентов блокированы абсолютно все моторные функции, все внешние проявления психики. Оказалось, что у больной есть сознание и что она способна адекватно отвечать на вопросы, задаваемые экспериментатором. Но эта коммуникация устанавливается на основе фиксации и расшифровки ее мозговых процессов с помощью функционально-магнитной резонансной томографии (6).

На горизонте расшифровка мозговых кодов процессов мышления. Достижение этой (пока еще не слишком близкой) цели способно

изменить традиционные типы человеческих коммуникаций, нарушить фундаментальный принцип относительной закрытости субъективного мира личности, на котором базируется наша социальная самоорганизация. Относительная закрытость субъективного мира означает, что личность открывает его по своей воле, как правило, частично и дозированно в зависимости от определенного коммуниканта, что это включает феномены обмана (в том числе защитного, добродетельного обмана), сокрытия информации и различные виды дезинформации. Проведите мысленный эксперимент и постарайтесь представить, что произошло бы с нами, с социальными институтами, экономическими структурами, партиями, государственными органами, если вдруг все начали бы говорить правду и только правду? Никто никому не врет! По-видимому, при нынешнем уровне массового сознания и социальной организации это вызвало бы неимоверный хаос во всей общественной жизни.

Перспектива расшифровки мозговых кодов мышления ставит острые вопросы. Ведь для расшифровки кодов нужна сложная технология. Кто будет владеть ею, в чьих руках она может оказаться? Кто останется «закрытым», кто кого будет «открывать» и зачем? Все это ставит принципиально новые проблемы развития земной цивилизации, в том числе связанные с углубляющимся антропологическим кризисом и антропотехнологическими преобразованиями человека.

Здесь нет места для их обсуждения (на эту тему опубликовано немало работ. См., напр.: [7]). Ясно одно, что расшифровка мозговых кодов явлений сознания открывает еще более масштабные перспективы преобразования человека, чем расшифровка генетического кода, но вместе с тем она влечет и еще большие риски и угрозы, еще более трудные проблемы.

Хочу остановиться только на одном сюжете, связанном с вопросами преобразования сознания. Это – в высшей степени актуальная сейчас проблема безопасности, касающаяся не только нашей страны, но человеческой цивилизации в целом, существованию которой все больше угрожают раздоры и войны.

Усиливается убеждение, что создание все более мощных и совершенных военных технологий уничтожения живой силы и разрушения инфраструктуры противника само по себе уже не способно обеспечить условия безопасности нашей страны и может иметь эффект бумеранга, ибо такие же технологии способен применить против нас и противник. Все это в равной мере относится и к другим развитым странам.

Возникает необходимость нового стратегического вектора в развитии технологий обеспечения безопасности. Их целью является создание эффективных средств воздействия на сознание, психику противника, подавления его агрессивности и злой воли.

О таких средствах прямого воздействия на сознание не принято говорить. Либералы наши сразу поднимут «бурю негодования», не говоря уже о западных рыцарях пресловутой политкорректности, прикрывающих этим жупелом свое личностное ничтожество, трусость, двуличие, страх потерять свое комфортное существование.

Да, публично говорить нельзя. Но хорошо известно, что в «ДАРПе», в других закрытых структурах США и наверняка в ряде других государств этим усиленно занимаются (не знаю, как у нас; если мы здесь сильно отстаем, то это – беда).

Перед нами огромная масса одержимых идеей Мирового халифата. Их число неуклонно растет. Они уже обсемили Европу с ее самоубийственной миграционной политикой. Что с ними делать? Завтра выйдут навстречу тысячи живых бомб. Как с ними справиться? Я не говорю уже о тех, кто без бомбы, но, тайно, с ножом.

Это реальная грозная проблема, и прежде всего это проблема сознания. Она должна стать предметом основательных размышлений и, главное, решительных действий в плане создания эффективных защитных средств на основе достижений нейротехнологий и информационных технологий. Для этого, по моему убеждению, есть все реальные возможности.

В заключение я хотел бы кратко сказать о том глобальном контексте, в котором разворачиваются исследования сознания, мозга, искусственного интеллекта, используются их результаты и результаты всей системы НБИКС-конвергенции, подчеркнуть приоритетную роль в этом социогуманитарной составляющей НБИКС.

Несомненно, что проблема сознания сейчас образует эпицентр всех глобальных проблем человечества. Мы вошли в полосу глубокого антропологического кризиса, который связан с нарастанием абсурда в развитии нашей потребительской цивилизации, с ее неспособностью разорвать параноидальный круг: еще больше производить, чтобы еще больше потреблять, чтобы еще больше производить.... Это влечет неизбежную борьбу за ресурсы, дальнейшее обострение экологических, энергетических, демографических, экономических кризисов глобального масштаба, дальнейшее обострение социальных, межгосударственных, межрегиональных, межконфессиональных конфликтов.

Общая перспектива нашей цивилизации выглядит крайне мрачно, о чем много написано и наговорено на разных континентах, однако не видно никаких обнадеживающих результатов совместных противодействий. Политические элиты погружены в свои вязкие текущие проблемы, в свои профессиональные игры между собой и с избирателями – им некогда думать, что будет через 20-30 лет. Интеллектуальная элита то и дело с тревогой говорит об этом, но чаще предпочитает вытеснять, отодвигать на задний план сознания грозные судьбоносные проблемы земной цивилизации. Пока не видно той силы, которая была бы способна что-то решительно изменить.

Вот почему главной философской проблемой нашего времени является экзистенциальная проблема сознания – о подлинных смыслах жизни и деятельности, о воле и дряблости духа, о судьбах земной цивилизации. Но я убежден, что философ, несмотря ни на что, должен поддерживать оптимистическую перспективу, крепить силу духа и веру в благие творческие силы разума, способного находить решения глобальных проблем, пути выхода нашей цивилизации на новый уровень развития.

ЛИТЕРАТУРА

1. *Риццолатти Джакомо, Синигалья Корrado*. Зеркала в мозге. О механизмах совместного действия и переживания. М.: Языки славянских культур, 2012. – 205 с.
2. *Иваницкий А. М.* «Чтение мозга»: Достижения, перспективы и этические проблемы // Журн. высш. нервн. деят., 2012. Т. 62, # 2. С. 1 – 10. См. также: *Дубровский Д. И.* Явления сознания и мозг: проблема расшифровки их нейродинамических кодов (Доклад на научной сессии Общего собрания Российской академии наук «Мозг: Фундаментальные и прикладные проблемы», 15 декабря 2009 г.) // Мозг. Фундаментальные и прикладные проблемы. М.: Наука, 2010.
3. *Metzinger Thomas*. The Ego Tunnel. The Science of the Mind and the Myth of the Self. Basic Books, N.-Y., 2009.
4. *Tononi G.* The Information Integration Theory of Consciousness // The Blackwell Companion to Consciousness / Ed. By Max Velmans and Susan Schneider. Blackwell Publishing Ltd., Oxford, 2007.
5. *Дубровский Д. И.* Проблема «Сознание и мозг»: Теоретическое решение. М.: Канон+, 2015. – 208 с.
6. *Оуэн А. М.* с соавт. Как обнаружить признаки сознания у пациентов в вегетативном состоянии // Горизонты когнитивной психологии. М.: «Языки славянских культур», 2012.
7. *Глобальное будущее. Конвергентные технологии (НБИКС) и трансгуманистическая эволюция*. Под ред. Д. И. Дубровского. М.: Издательство МБА, 2013. – 272 с.

ИСКУССТВЕННЫЕ ОБЩЕСТВА ARTIFICIAL SOCIETIES

马卡落扶 «人工社会»

Макаров Валерий Леонидович

*академик РАН, сопредседатель НСММИ РАН,
доктор физико-математических наук,
профессор, директор Центрального экономико-
математического института РАН, г. Москва*

Makarov Valery L.

*academician of the Russian Academy of Sciences,
Doctor of Science (Physical and Mathematical),
professor, Director of the Central Economics and
Mathematics Institute of the Russian Academy of
Sciences, Moscow*



Аннотация. Рассматривается направление исследований, которое занимается компьютерным моделированием социальных процессов. Искусственные общества представляют собой многоагентные системы. Агенты обладают сложным поведением, их взаимодействие формирует определенные устойчивые структуры – искусственные социальные системы. Обозначаются методологические и эпистемологические проблемы «искусственных обществ».

Ключевые слова: компьютерные модели социокультурных процессов, мультиагентные системы, искусственные общества.

Annotation. The direction of research that deals with computer modeling of social processes is considered. Artificial societies are multiagent systems. Agents have complex behavior, their interaction forms certain stable structures – artificial social systems. The methodological and epistemological problems of «artificial societies» are indicated.

Keywords: computer models of sociocultural processes, multiagent systems, artificial societies.

Фридрих Ницше, правда, не очень понятно, сказал примерно следующее: «Если Вы хотите понять что-то лучше, детально, до конца, попробуйте это сконструировать. И если Вы это сконструируете, значит, Вы поняли». Во времена Ницше, конечно, не было компьютеров, но как они появились, сконструировать стало легче, особенно что-то сложное. Мы сегодня обсуждаем искусственный интеллект. «Искусственный» – значит, мы пытаемся сконструировать «интеллект». И конструируем его с помощью компьютеров.

Когда они появились, возникло мощное средство для этого. А человеческое общество не менее сложная конструкция, чем отдельный человек. Естественно, возникла идея конструировать искусственное общество. И эта мысль настолько захватила всех, что возникло целое движение, и тысячи людей сейчас занимаются этими вопросами. Появился *Journal of Artificial Societies and Social Simulation*, JASSS («Журнал искусственных обществ и имитационного моделирования общества»), где печатают статьи в основном об искусственных обществах. Мы у себя в институте тоже сейчас создали журнал под названием «Искусственные общества», который приобретает всё большую популярность, и тут огромное поле разнообразной деятельности, хотя началось всё давно. Часто цитируемый Джон фон Нейман, который создал теорию автоматов, предположил, что автоматы, которые сами себя могут производить, тоже образуют общество, общество автоматов. И это тоже пример искусственного общества. И, естественно, нам хочется иметь модели, похожие на настоящие общества.

Благодаря суперкомпьютерам стало возможно создавать более сложное искусственное общество с большим количеством агентов. Вместо людей сейчас говорят агенты, ведь агентами могут быть не только люди. Например, в нашем институте одна девушка занимается песцами, обществами песцов и их тоже называет агентами. Появляется термин «многоагентные модели». Искусственные общества появились не только у ученых, например, сейчас есть популярная игра «*Second Life*», где можно конструировать другую жизнь, и это тоже искусственное общество, и ее сейчас тоже пытаются понять ученые. У меня есть знакомая, крупная бизнес-леди, хотя она доктор наук, она тоже играет в «*Second Life*». В жизни она очень комплексует из-за отсутствия слуха, а в «*Second Life*» она великая певица. То есть искусственные общества создаются народом, а не только учеными. Рекорд пока у американцев, они всё американское население, 300 с

лишним миллионов людей заложили в эту модель общества, но оно исключительно примитивно, все люди копии друг друга. А известен этот проект тем, что в этом обществе очень хорошо моделируется распространение разного рода болезней, например, грипп, с какой скоростью какие штаты захватываются. А потом они подставили вместо болезни слухи, и эта конструкция оказалась куда сложнее модели гриппа или Эболы. Мы тоже разработали в своем институте на суперкомпьютере подобную демографическую модель, где всё население России 140 миллионов, но тоже это некоторый обман, это люди-копии, разных агентов там 15 - 20. Это, конечно, обидно для людей, что они являются копиями друг друга. Прогонка была сначала начата на «Чебышеве», сейчас на «Ломоносове». Это очень трудоемкое дело, исследовать, как распространяются по территории России некоторые явления. У нас получается, что всё большая часть людей в модели перемещаются в европейскую часть и теплые места, они не любят холод, Сибирь и Дальний Восток. Сейчас есть попытки изучить распространение мнений, в частности, религии. Есть ли шансы, что укрепится православие? Но эти исследования только начинаются. А модель сама по себе интересна тем, что с ее помощью можно промоделировать миграцию.

В искусственных обществах это все числовые компьютерные эксперименты, и вопрос в том, имеет ли это отношение к действительности. Сейчас прошел пятый мировой конгресс по искусственному обществу и социальному моделированию. Там был один доклад, где в обществе существовало три пола. Автор придумал способы реализации, но все его начали резко критиковать. Какое отношение это имеет к жизни? На что получили ответ, что это просто эксперимент. И такая ситуация – одна из главных проблем всех, кто занимается компьютерным моделированием.

Вы знаете, что у математиков главный результат – это доказанная теорема. Например, в рыночной экономике математики доказывают модель рыночного равновесия, где уравнивается спрос и предложение, известную теорему о неподвижной точке цены. А есть проблема, как формируются не цены, а социальные нормы. И их не меньше, чем цен. Проблема того, как нормы формируются, их динамика, и тут не получается математической модели. А в моделях искусственных обществ это получается. В JASSe есть много экспериментов, которые показывают, как формируются социальные нормы. От самых простых, что нельзя сморкаться в занавеску, до норм типа необходимости

уступать место даме. Но в Америке, например, на это могут обидеться. Таким образом, социальные нормы исчезают. Математические модели тут не получаются. Есть цитируемая работа касательно нормы, сколько надо давать на чай. Критика в том, что это не теорема, вы показали, что норма формируется, а я сейчас возьму другие данные и правила, и всё будет наоборот. Это не результат. Вопрос, что вообще здесь является результатом? JASSS сейчас просит всех давать не только саму работу, но и приложение с программой, как это проводилось. Каждый может проверить, нравится – повтори. У физиков это принято, открытие признается только тогда, когда другая независимая лаборатория повторила эксперимент. И тем самым происходит шаг вперед: результаты компьютерных экспериментов – это результат, который должен признаваться. Виктор Истратов написал большую статью о социальных нормах, как их понимают психологи, компьютерщики и пр. Есть нормы, живущие столетиями и тысячелетиями, а есть такие, которых и через 10 лет уже нет. Это очень интересная сфера, и если вы построите свое искусственное общество, то на нем можете проводить любые эксперименты. Выступавший передо мной упомянул об обществе с роботами, и такие модели уже создаются.

Есть движение *big data*, где обрабатывается и анализируется огромное количество *messages*. И выясняется, что чуть ли не 80 % сообщений сделали боты, а не люди. То есть нужно отличать, когда робот сказал что-то в Интернете, а когда человек. Хочешь или не хочешь, наше общество идет к моменту, когда с роботами придется общаться.

Надо, чтобы больше людей и в России занимались таким моделированием социальных явлений. На «Ломоносове», к сожалению, мало специалистов в области общественных наук. А их участие в суперкомпьютерных проектах важно и необходимо.

ПОСТНЕКЛАССИЧЕСКАЯ РАЦИОНАЛЬНОСТЬ И ИНФОРМАЦИОННОЕ ОБЩЕСТВО POST-NONCLASSICAL RATIONALITY AND THE INFORMATION SOCIETY

斯切兵 《合理性和信息社会》

Степин Вячеслав Семенович

*академик РАН, доктор философских наук,
профессор, г. Москва*

Stepin Vyacheslav S.

*academician of the Russian Academy of Sciences,
Doctor of Science (Philosophy), Professor, Moscow*



Аннотация. Проблемы искусственного интеллекта рассматриваются в контексте парадигмы постнеклассической рациональности. Возникает особый тип взаимодействия наук — естественных, технических, гуманитарных и социальных. Это целостный комплекс саморазвивающихся естественных и искусственных систем.

Ключевые слова: искусственный интеллект, постнеклассическая рациональность, саморазвивающаяся система.

Annotation. The problems of artificial intelligence are considered in the context of the paradigm of post-non-classical rationality. There arises a special type of interaction of sciences – natural, technical, humanitarian and social. This is a holistic complex of self-developing natural and artificial systems.

Keywords: artificial intelligence, post-nonclassical rationality, self-developing system.

Я буду говорить о постнеклассической рациональности и попытаюсь связать это с проблемами, которые сейчас в технике, технологиях и вообще в культуре доминируют. Если рассматривать историю науки Нового времени, а для меня наука, если рассматривать ее вместе с возникновением естествознания, в общем-то, начинается в Новое время. До этого, в предшествующие периоды, она тоже, конечно же, была, но соединение математики с экспериментом — это галилеевская программа — породило физику и механику, и это был первый тип науки, который стал доминировать и оказал влияние потом на другие области естественнонаучного знания, на социальные науки. Это был первый тип рациональности, который часто обозначают как классическую рациональность. Затем возникла эпоха, когда в науке произошли серьезные преобразования, часто обозначаемые как неклассическая рациональность, связанные с научными революциями в естествознании начала двадцатого столетия, прежде всего, с появлением квантовой релятивистской физики. Но сейчас в науке и технологиях происходят новые революционные сдвиги, которые можно обозначить как новый тип рациональности, назвать их можно постнеклассическими. Вообще, я пишу на эту тему уже давно, первая моя книжка — «Научные революции и динамика культуры» — это книга 1987 года, издана была в Белоруссии, где я еще работал, в этот год как раз переехал в Москву, стал директором института естествознания и техники. Так вот, я дал там дескрипцию этого понятия, а в 1987 году или в 1989 году в «Вопросах философии» была статья, где уже термин применялся.

Возникает вопрос, в чем критерий различий этих типов рациональности? Как их различить? Первое различие, которое здесь проводится, — различие по способам объяснения, описания, идеалам и нормам науки, а также по типу рефлексии над знанием, прежде всего, философской рефлексии. Немного в тени оставался вопрос о том, что за тип объектов позволяет осваивать каждый из этих типов рациональности. Дальнейшая работа привела к тому, что каждый тип рациональности оказался связанным с особым видением объектов, которые изучает наука, и с особым типом научной картины мира, которая тоже строится как системность, как системное видение, в котором неявно заложены представления об особом типе объекта. Эти типы объектов можно обозначить таким образом: это простые системы, в основном механические; они связаны с классическим типом

рациональности, являясь предметом исследований этого типа, и все идеалы и нормы, тип философской рефлексии приспособлены к тому, чтобы дать знание об этом типе объектов. Второй тип объектов — объекты саморегуляции; это сложные гомеостатические системы саморегуляции. С ними столкнулась, не видя их структуры, описывая их, скорее, феноменологически, вначале квантовая релятивистская физика. Потом эта революция продолжалась: в сороковых годах это работы Винера, Розенблата — то, что потом породило первый цикл системных исследований. В шестидесятых годах появились системные исследования Берталанфи и других авторов — это изучение, по существу, сложных систем под углом их саморегуляции, гомеостаза. И, наконец, третий тип систем или объектов — сложные саморазвивающиеся системы, переходящие от одного типа гомеостаза к более сложному типу гомеостаза. Концепция сложных развивающихся систем — то, что требует особого типа рациональности, рефлексии, понимания, объяснения, описания и так далее, то есть новых идеалов и норм науки.

Каково отличие этих систем? В шестидесятых годах был хороший известный систематик — Гелий Поваров, который предложил идею системной сложности техники в историческом развитии цивилизации. Он выделил, начиная с первобытного строя и до наших дней, несколько типов системной сложности техники, когда поначалу это техника простых предметов (палеолит). Затем техника сложных составных предметов — это, в основном, первые городские и сельские цивилизации древности. Далее техника простых систем — это уже начало цивилизаций, которые я обозначаю как техногенные. Затем — техника саморегулирующихся сложных систем. Понятие «сложная система» включает в себя два смысла, которые часто неразличимы. Сейчас синергетика очень много говорит о саморегулирующихся системах, то есть о самоорганизующихся. Понятие самоорганизации относится к любым сложным системам, но сами сложные системы надо разделить по крайней мере на два класса: саморегулирующиеся, гомеостатические системы и системы с развитием, эволюцией, где есть регуляция особого типа. Различая эти вещи, мы можем задать вопрос: какой категориальной сетки требует понимание и описание этих систем? Оказывается, что любое понимание системного объекта предполагает следующие категориальные смыслы:

1. Соотношение части и целого, вещи и процесса — это как бы субстрат системы;

2. Тип взаимодействия, где ключевым понятием является причинность, то, как она понимается;

3. Пространство и время.

Кстати, вот в этих трех позициях строится всегда научная картина мира: субстрат, взаимодействие, пространство и время.

В понимании субстрата простых систем, понимании части и целого полагается, что свойства частей исчерпывающе определяют свойства целого. Целое образуется благодаря свойствам входящих в него частей. Идея какого-то особого системного качества не присутствует в описаниях в классике, и вся механика, и классическая термодинамика, и классическая термодинамика Максвелла, Вебера, и все предшествующие основываются на такого рода представлениях. Далее предполагается о вещи и процессе, что вещь — нечто первичное, а процесс — взаимодействие одной вещи с другой, силовое воздействие одной вещи на другую, которое меняет состояние вещей. Следующий момент: причинность. Для того, чтобы описать простые системы, достаточно свести причинность к лапласовской детерминации. Это строгая трактовка причинности, где, зная начальное состояние и силы, действующие на системы, мы можем предсказать со сколь угодно точностью её будущее. Это известная лапласовская идея, что, если бы я знал и мог выделить все атомы Вселенной и указать все действующие на них силы, я мог бы состояние Вселенной описать со сколь угодно точностью на сколь угодно дальнее время.

И, наконец, пространство-время. Пространство — это внешнее по отношению к системе, то, где разыгрывается взаимодействие объектов с течением времени; это абсолютные пространство и время Ньютона, которые он определяет как чувствилище Бога. Этого достаточно для выделения и описания простых механических систем. Кстати, в этой категориальной сетке описываются не только физические объекты, объекты механики, которые тогда доминировали в науке, под этим углом зрения рассматриваются и биологические процессы, и социальные процессы. Вспомним, что Огюст Конт называл свою социологию социальной механикой, и на первых порах полагалось, что можно, выделив в качестве атомов отдельных индивидов (это была идея Шарля Фурье), а в качестве их свойств определив их устремления, ценности, особые качества, понять, как будет развиваться общество. Фурье даже говорил, что можно найти какой-то универсальный закон жизни вроде закона всемирного тяготения, который описывал бы взаимодействие простейших элементов и

предсказывал, как будет развиваться состояние общества. Правда, Огюст Кот столкнулся потом с трудностями реализации этой программы: он один из первых сказал, что общество — это целостный организм, где есть особые качества, не сводимые к свойствам частей, и это уже было началом перехода к новому типу объекта, а значит, к новому типу рациональности.

Так вот, новый тип рациональности возникает, когда мы сталкиваемся с объектами, которые устроены как сложные, саморегулирующиеся системы, в этих объектах идут сложные стохастические взаимодействия, есть блок информации и управления, есть прямые и обратные связи, есть программа, в соответствии с которой система себя воспроизводит. Такого рода системы требуют создания новой категориальной сетки. Во-первых, часть и целое: целое не редуцируется к свойствам частей; всегда фиксируется некое особое системное качество, которое возникает благодаря взаимодействию частей. Ни одна часть в их свойствах или даже совокупность их свойств не описывает целого, то есть еще нужно представление о некоем системном качестве. Во-вторых, изменяется понимание вещи и процесса: в саморегулирующихся системах гомеостаз и есть процесс, он воспроизводит себя, а воспроизводство себя как процесса и есть вещь. Любой организм как предмет существует только благодаря тому, что он воспроизводит себя. Кстати, сам процесс воспроизводства клетки, органов, организма благодаря обмену веществ с внешней средой происходит не как слепое копирование и повторение — там всегда накапливаются какие-то ошибки, и сейчас показано, что этот процесс всегда связан с элементами динамического хаоса, которые описываются в языке синергетики (это работы Иваницкого, Еськова). Сейчас это на переднем плане описания процесса гомеостаза с учетом элементов динамического хаоса, которые обязательно присутствуют при воспроизводстве системы. Здесь вещь и процесс меняются местами: процесс становится первичен. Значительно расширяется и понимание причинности: лапласовская причинность сохраняется, но она дополняется так называемой вероятностной причинностью, описывающей стохастические взаимодействия в подсистемах. Эта стохастическая вероятностная причинность выступает не как вторичная, а как первичная. Стохастические взаимодействия знала и классика, но классика считала, что их можно свести к лапласовской детерминации; современные представления говорят, что нельзя. Одним из первых зафиксировал необходимость такого понимания

причинности Нильс Бор. Знаменитые дискуссии на Сольвеевских конгрессах с Эйнштейном, где Эйнштейн отстаивал классическое понимание причин, а Бор показывал, что имеются квантовые скачки, что есть случайности, лежащие в фундаменте мироздания, а потому вероятностная причинность должна быть взята как базис. И, наконец, пространство и время. Выясняется, что в системах саморегуляции, системах гомеостатического характера возникают внутренние пространство и время, отличные от внешнего физического пространства. В биологии это известно: экологические ниши и ареалы обитания живого, а также внутреннее время организма, биологические часы. В социальных процессах происходит то же самое: есть социальное пространство и социальное время, отличные от внешних природных пространства и времени.

Что происходит с категориальной сеткой для саморазвивающихся систем? Сохраняется идея системного качества, не сводимого к свойствам элементов. Но, поскольку система развивается, в процессе развития возникает новый уровень организации элементов. Принципиальная характеристика саморазвивающихся систем: они способны наращивать новые уровни своей организации, причем каждый новый уровень организации, возникающий в процессе эволюции, начинает воздействовать на нижележащие уровни. Это принципиально важная методологическая идея: он оказывает на них свое воздействие и ограничивает некоторые процессы, функционирование элементов, их свойства на низшем уровне. То есть параметры порядка управляющей системы всегда лежат на высшем уровне, а не внизу. Они возникают в процессе наращивания системой новых уровней своей организации, и каждый новый уровень всю систему переделывает, накладывает некоторое ограничение на функционирование предшествующих уровней — и только так система превращается в новое целое, в новый гомеостаз. Значит, переход от одного типа гомеостаза к более сложному совершается за счет появления новых уровней организации. Итак, пока система развивается, пока идет эволюция, в системе всегда будут наращиваться уровни организации. Конечно, система может быть разрушена в этом процессе, она может и деградировать, но, если она выживает и усложняется, там постоянно усложняется иерархия. И еще раз повторю: это свойство высших уровней воздействовать на низшие уровни — ключевое свойство саморазвивающихся систем.

С этим наука сталкивается на каждом шагу. Это было зафиксировано уже в рамках физики. Например, как застраивается электронная оболочка атомов: работает принцип Паули, два электрона в одном состоянии не могут быть на электронной оболочке, сразу налагается ограничение. Если у вас «электронный галстук», там могут быть разные электроны в разном состоянии, но как только они вступают во взаимодействие и конструируется сложная система, сразу налагается ограничение по принципу Паули. Когда строится ядро, протоны и нейтроны внутри ядра образуют новое состояние нуклона, где протон переходит в нейтрон и наоборот, там есть дефект массы — это известная вещь. Сумма массы протонов и нейтронов вне ядра не равна массе нуклона, то есть сумме масс протона и нейтрона внутри ядра. Это простые примеры, которые давно известны. Известны и примеры о функционировании клеток — более сложные вещи. Когда они входят в органы более сложного целого и когда они развиваются в органы более сложного целого, накладываются свои ограничения. В свое время (что было лично для меня удивительно, поскольку я считал, что неклассическое мышление начинается с квантовой релятивистской физики, а оказалось, что это не совсем так) в истории медицины был такой период, связанный с работами великого физиолога Бернара, патологоанатома Конгейма, которые зафиксировали, что организм нужно рассматривать как процесс. Тогда в медицине была ходячая «механическая» метафора об организме как о паровой машине: он поглощает пищу, как паровая машина сжигает уголь, за счет чего идут все жизненные процессы, осуществляется движение мышц и так далее. Бернар и Конгейм зафиксировали экспериментально доказанные вещи, что это не так, что пища перерабатывается, возникают новые части организма, жиры, которые потом перерабатывает печень. Получается, что организм — это такая паровая машина, которая сжигает свои собственные детали, собственный материал в процессе своей жизненной деятельности. То есть организм — это процесс. У Бернара была еще одна интересная идея: он говорил, что каждая клетка, когда она входит в орган (а клетка — живой организм), приспособляется к другим клеткам, и поэтому обретает особые свойства, которые заставляют ее быть клеткой кожи, мышц, волосяного покрова и так далее, то есть она специализируется. Эту идею забыли, но потом она была восстановлена уже сейчас с открытием стволовых клеток и исследованием, например, такого процесса, который показывают эксперименты, — морфогенеза. В ходе него идет развитие

эмбриона, у головастика в начальной стадии, когда уже началось разделение головной части и туловища, берут из туловища клетку, которая в принципе могла бы развиться в кишечник, вставляют ее в головную часть, и эта клетка развивается в глаз, становится глазом, допустим, лягушки. То есть клетка в зависимости от клеточного окружения может приобретать разные свойства. Теперь выяснили, что обычные клетки могут превратиться в стволовые. Клетка содержит в себе полный код организма, она может стать любым органом, но клеточное окружение, возможные химические ингибиторы и электрические воздействия (это сейчас исследуют) блокируют части кода и заставляют клетку развиваться в особые органы. Таким образом, это процессуальная система.

Так вот, когда мы рассматриваем современные технологические процессы, можно сказать, что сейчас вся техническая революция современности основана на переходе к сложным системам, в том числе, системам саморегуляции и саморазвития. Система саморазвития становятся передним планом технологического и научного освоения. И какова тогда их категориальная сетка? Для того, чтобы их понять, надо сказать, во-первых, что системные качества меняются при появлении новых уровней организации системы. Возникают системные качества целой системы, поскольку она эволюционирует. Во-вторых, процессуальность надо понимать в двух смыслах: в смысле воспроизводства устойчивого состояния системы и в смысле нарушения этого устойчивого состояния, то есть в виде фазовых переходов. Кстати, эти фазовые переходы описываются сейчас в синкретических терминах; это системный переход, связанный с динамическим хаосом. И здесь есть свои описания: кооперативные эффекты, точки бифуркации, наличие множества сценариев развития в точках бифуркации, реализация одного из сценариев, идея русла, по которому может идти система, и, наконец, идея режима с обострением, которая наиболее исследована в работах Курдюмова. Режим с обострением завершает эволюцию на этом этапе фазового перехода, и появляется новый уровень организации, который будет менять предшествующие уровни и воздействовать на них.

Если рассматривать эту категориальную сетку, следует сказать, что со временем меняются системные качества систем с саморазвитием, меняется понимание процесса (есть два вида понимания процесса: один процесс — воспроизводство, а другой — мутация, изменение воспроизводства и возникновение нового порядка, идея порядка из

хаоса). Относительно причинности возникают интересные вещи: сохраняется лапласовская причинность, существует причинность вероятностная, и возникает еще целевая причинность. Это когда появляется странный аттрактор, который набирает силу и определяет русло развития, то есть определенный сценарий, в то время как остальные сценарии теряют свою вероятность, меняется статистика, теория вероятности должна быть другой. У Колмогорова была такая идея, что нужно разработать теорию вероятности с изменением вероятностных мер. Для систем с саморазвитием это обязательное условие, потому что как только начинает реализовываться определенный сценарий развития, остальные «выгорают» (это термин Курдюмова), то есть они понижают свою вероятность. Возникает такая идея, что система идет к какому-то исходному своему состоянию, целому, завершающему этапу, к новому уровню организации, и это может быть описано в терминах целевой причинности. Это не значит, что система обладает целеполаганием; это естественный процесс движения странного аттрактора. И, наконец, пространство и время. Мало сказать, что есть внутреннее пространство-время системы, придется сказать, что они меняются по мере эволюции. Это блестяще описал в свое время Вернадский, когда он говорил, что геохимические процессы Земли до возникновения жизни идут сотни миллионов и миллиардов лет, но, когда возникает жизнь, они ускоряются, и геохимия переходит в биохимию. Все те же процессы совершаются в совершенно ином ритме и с другими скоростями, то есть пространство и время системы меняются. Таково должно быть понимание систем с саморазвитием.

Какие здесь возникают проблемы? Когда мы так рассматриваем три типа систем и три типа рациональности, мы также видим, что в зависимости от решаемых наукой задач они соседствуют, наука становится гетерогенной. Каждый новый тип не уничтожает предыдущего, он соседствует с ним, но ограничивается в своих претензиях и возможностях. Если это так, то существует ли преемственность между этими типами рациональности? Эта проблема решается в моей последней работе следующим образом. Я ставлю вопрос так: что такое гомеостаз? Это устойчивое состояние саморазвивающейся системы, это ее фрагмент. А что такое простая механическая система? Это фиксация того, что, когда есть устойчивое состояние, наличествуют параметры порядка, и, если я рассмотрю эти параметры порядка как устойчивые и воспроизводящиеся, пока устойчива система, они существуют. Как говорится, во времени что вперед, что назад — они

одни и те же. И тогда мне не нужна эта идея стрелы времени, я могу описывать каждую вещь, взаимодействие вещей и каждый процесс с позиций устойчивых параметров порядка, как это делается в механике. Если я хочу, например, описать, как Земля движется вокруг Солнца, мне не нужно воспринимать Солнце как процесс. Оно действительно процесс, но я от этого абстрагируюсь, даже если я это знаю. Вот сейчас я знаю, что там есть хромосфера, фотосфера, идут ядерные реакции — это все кипит и бурлит, но я могу рассмотреть это просто как массу, которая имеет определенное значение, — вот вам системный параметр. Это, кстати, важный параметр, ибо если масса теряется, то Солнце превратится в планету, остынет и уже не будет звездой. А планета Земля? Это ведь тоже процесс взаимодействий биосферы, геосферы, литосферы, атмосферы, человеческая деятельность — все ускоряется, все меняется, как описывал Вернадский. Это процессуальная система, но я абстрагируюсь и говорю, что это тоже масса, и теперь мне нужны только две массы и расстояние между ними; и я с точностью могу описать орбиту земли и других планет, абстрагируюсь от всех процессуальных вещей, рассматривая их [планеты] как идеализированные неизменные ограниченные тела, могу предсказывать движение по орбитам, солнечные и лунные затмения — и все будет сходиться. Вот вам и механическая система. Что же тогда относительно преимущества? Я отвечу так: первичной является саморазвивающаяся система. С самого начала возникновения метagalактики идут процессы, с момента большого взрыва идет эволюция, но мы не видели этого. Если и видели, то описывали в каких-то терминах, которые были у нас под рукой. Мы начали с фиксации неких системных параметров, зафиксировали это, затем пришли к гомеостазу, начали изучать гомеостаз как изменяющийся в другой гомеостаз. Тогда вы мне скажете: а была ли эволюция раньше, до этого представления о сложных системах? Да, была. У Ламарка, например, была, но он описывал ее механически; он говорил так: упражнения создают орган, чем больше я упражняю мышцу, тем больше она разрастается; есть флюиды нервной системы, которые идут от органа к органу и реально меняют сами органы. Упражнение создает орган, новые органы — новые виды. Вот механическое объяснение; оно, правда, ушло, но идея эволюции Ламарка осталась, мы его называем великим эволюционистом, но он использовал механическую систему и всю категориальную сетку, с точки зрения которой он и описывал эволюцию.

Теперь эволюцию нужно описывать в ином понимании. И только тогда самое сложное — саморазвитие — есть исходное целое, фрагменты и аспекты которого могут описываться при решении других познавательных задач в других областях знания, и в этом смысле вы сохраняете и классическую, и неоклассическую рациональность, хотя постнеклассическая будет доминировать. Это первый момент.

Второй момент: в постнеклассическом подходе, когда я вижу саморазвивающиеся системы, радикально меняется проблема естественного и искусственного. Когда я говорю о том, что у меня в точке бифуркации возникает несколько сценариев развития, моя деятельность может активизировать один из сценариев, и тогда то, что я создам, будет искусственным, но, с точки зрения развивающейся системы, — это так же естественно, как и любой другой сценарий. Природа не реализовала всего того, что могла бы реализовать. Но, создав человека, она запускает через него новые линии своей эволюции. Так решается эта проблема: искусственное и естественное неразличимы с точки зрения саморазвивающейся системы. Далее, включение деятельности как компонента в саму систему — это включение самого человека, а раз человек включается в любую природную или социальную, любую систему как деятельный, тогда он становится частью системы. Эксперименты с системой, изучение системы накладываясь на это ограничение, из-за чего возникают социально-этические экспертизы научных исследований и проектов, возникает особый тип взаимодействия наук — естественных, технических и социальных; они должны быть целостным комплексом.

КОГНИТИВНЫЕ ВЫЧИСЛЕНИЯ НА ОСНОВЕ НЕЙРОННЫХ ГИПЕРСЕТЕЙ COGNITIVE CALCULATIONS BASED ON NEURAL HYPERNETWORKS

阿诺轻 «基于神经超网络的认知计算»

**Анохин Константин
Владимирович**

член-корреспондент РАН, доктор медицинских наук, руководитель отдела нейронаук НИЦ «Курчатовский институт», руководитель лаборатории нейробиологии памяти НИИ нормальной физиологии им. П. К. Анохина, г. Москва

Anokhin Konstantin V.

Corresponding Member of the Russian Academy of Sciences, Doctor of Medical Sciences, Head of the Department of Neuroscience of the Research Center «Kurchatov Institute», Head of the Laboratory of Neurobiology of the Memory of the Research Institute of Normal Physiology named after P. K. Anokhin, Moscow



Аннотация. Обозначены этапы развития когнитивной науки. Сегодня идет подготовка третьей когнитивной революции. Она произойдет с созданием теории мозга. Для этих целей перспективна теория когнитивной гипернейронной сети.

Ключевые слова: когнитивная революция, коннектом, когнитом, гипернейронная сеть.

Annotation. The stages of the development of cognitive science are indicated. Today the third cognitive revolution is preparing. It will happen with the creation of a theory of the brain. For these purposes, the theory of cognitive is promising as a hyperneuron network.

Keywords: cognitive revolution, connectivity, cognitive, hyperneuron network.

Мой доклад содержит два ключевых термина: когнитивные вычисления и нейронные гиперсети. Я разберу их по очереди. Но прежде необходимо обозначить основные тезисы этого сообщения. Они следующие:

1. В области когнитивной науки нам нужна Третья когнитивная революция.

2. Нам нужны принципиально новые идеи для создания искусственных когнитивных систем.

Первой когнитивной революцией принято считать то, что Хомский считал когнитивной революцией XVII-XVIII века, Второй когнитивной революцией считается та, которая возникла в США в середине прошлого века и продолжается в виде активного развития когнитивных наук сейчас. Но этого недостаточно, поскольку у нас нет теории когнитивной науки.

Прошло много лет, и в этом году не один юбилей. 1936 год – фундаментальная работа Тьюринга, 1946 год – первая машина, созданная Дж. фон Нейманом, 1956 год – создание синтеза когнитивных наук и терминологии искусственного интеллекта (ИИ). Но надо понимать, что Тьюринг строил свою универсальную вычислительную машину, беря логику вычислений из логики собственного разума потому, что никакие серьезные вещи относительно того, каким образом мозг делает это, как он устроен и каковы его операции, в то время не были еще известны и пока не становятся известны и сейчас. Это только момент бурно развивающейся нейронауки. Дж. фон Нейман работал с электронно-лучевыми трубками для того, чтобы симулировать цифровые операции, и ему тоже было ничего неизвестно в 1946 году, так же, как и К. Шеннону, для того, чтобы понять, что реально и конкретно делает нервная система. Поэтому сегодня нужны новые принципы и подходы. Вообще, я думаю, мало кто будет возражать сегодня против этого тезиса – все говорят об этом в последние годы. Мне кажется, содержательная сторона состоит в следующем: это не только необходимо, но и возможно, и возможно настолько, что это должно стать предметом ежедневной работы. Эти два вопроса должны стать предметом ежедневной работы больших коллективов ученых, а это пока не так. Но разберем, что у нас есть в отношении такой работы и движении в эту сторону.

Когнитивные вычисления

Термин «когнитивные вычисления» серьезно был поднят компанией IBM (International Business Machines Corporation), которая каждый год проводит стратегические научные конференции, посвященные перспективам развития компании в разных направлениях. В 2006 году она провела в Альмадене – в своем исследовательском центре (IBM Almaden Research Center) – конференцию, которая называлась «Когнитивные вычисления». Еще некоторое время назад доклады приглашенных людей были на сайте IBM. И там можно увидеть, насколько серьезно были подобраны эти люди. Там были и нобелевские лауреаты, но вопрос не в статусе; там были люди, которые не носили никакого статуса, но весь опыт за последние 10 лет показал, что они были выбраны правильно. Это были и нейрофизиологи, и те, кто занимается созданием искусственного интеллекта.

IBM серьезно с этого момента перестроила свою стратегию. Они продали свои рутинные компьютерные компании (персональные компьютеры и ноутбуки отошли китайской компании «Lenovo») и взяли установку на развитие вычислений, которые делает мозг млекопитающих, поставив вначале задачу создать компьютеры, которые будут работать, как мозг крысы (с такой же эффективностью), потом кошки, обезьяны и, в конце концов, человека. Таким образом, основной лозунг этой конференции, ее окончательная и предельная цель – понять, как и когда мы можем механизировать когнитивные операции. То есть редуцировать их до механических процедур, воплощаемых в искусственных системах.

Приблизительно в это же время большая программа движения в сторону таких операций (того, что было названо нейроморфным ИИ) была начата DARPA (Defense Advanced Research Projects Agency). Его специалисты посчитали, что по сравнению с биологическими системами, интеллектуальные машины, которые существуют сейчас, менее эффективны в миллионы или даже в миллиарды раз. И ясно, что никакое линейное развитие существующих операций и архитектуры не способно решить это, особенно в условиях, когда биологическая нервная система сталкивается с комплексными задачами (так называемый универсальный или общий ИИ). Поэтому DARPA была поставлена задача перестать финансировать исследования по классическому экстенсивному развитию архитектуры ИИ и найти принципиально новые пути для создания машины не фон-неймановской

архитектуры, а работающих на принципах операций, производимых головным мозгом.

В результате этого DARPA и IBM в США объединились в создание программ, одна из которых получила название «Система нейроморфной адаптивной пластически шкалируемой электроники» или сокращенно SINAPS. Другая программа IBM – когнитивные вычисления с помощью синаптроники и суперкомпьютеров.

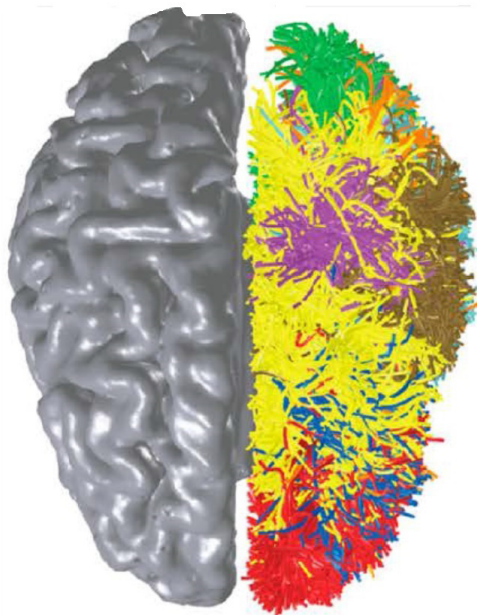
В чем же основные идеи и подходы, рассматривающие то, как работает мозг, в этих программах? Первая из этих идей состояла в том, чтобы заменить принципы фон-неймановской архитектуры с локализованной памятью. Это известная, так называемая проблема «бутылочного горлышка» архитектуры фон Неймана – шины пропускания памяти к памяти операций на системы с распределенной памятью, где память хранится и образуется в элементах всей сети пересечения различных элементов. И в качестве таких пластических элементов, имитирующих контакты в нервной системе, были взяты мемристоры.

Компания IBM в течение последних 10 лет очень активно развивает это направление. Были выпущены прототипы так называемых нейроморфных чипов, которые все время обновляются. Важно то, что они оперируют гораздо быстрее, чем существующие классические архитектуры.

Следующая важная вещь – это энергопотребление. IBM исходно ориентировалась на использование биологических принципов распределенных сетей для того, чтобы снизить энергопотребление. Энергопотребление операций, построенных на нейроморфных чипах, уже в тысячу раз меньше, чем на классических компьютерных архитектурах.

Другое направление IBM очень интересное. Специалисты IBM в исследовательском центре в Альмадене занялись исследованием того, как устроены архитектуры связей в головном мозге у млекопитающих, в частности, у приматов. Совершенно непростая и нетривиальная вещь для специалистов такой компании. На рисунке ниже – вырезка из статьи, которая была опубликована в докладах Нью-Йоркской академии наук и вышла от имени IBM «О сетевой архитектуре длиннодистантных путей в мозге макаки», где были прокартированы различные участки коры головного мозга, показанные здесь на коннектограмме различными модулями. И здесь видны разные доли: височная доля, затылочная доля коры связаны

друг с другом, если их редуцировать по так называемой сердцевине, где проходят основные когнитивные операции, которые вовлекают большинство из когнитивных процессов в мозге млекопитающего. Зачем это нужно? Почему недостаточно только распределенной памяти с мемристорными синапсами, на которых делаются чипы IBM? Это в значительной степени связано с тем, что необходимы поиски решения двух принципиальных проблем в области различия между искусственными когнитивными системами, существующими компьютерами и, естественно, мозгом.



Первое. Наш мозг оперирует с очень небольшой тактовой частотой (около 500 Гц). Это столько, сколько нужно между двумя спайками в обычных операциях клеток. Современные машины используют гигагерцевые частоты для процессинга. И, тем не менее, за 1 секунду человек способен распознавать сложнейшие вещи. Да, эти гигагерцевые машины распараллелены, как разбирал Хоукинс, однако это ничего не решает. Количество шагов, которые они делают, огромное, значит, нужны другие принципы. Также пока совершенно неясно, каким образом в новых архитектурах решать такую проблему:

принятие решений, извлечение хранящегося у человека опыта (а это огромные объемы опыта) занимает около 200-300 миллисекунд – это глобальный доступ ко всей памяти. Каким образом огромное количество степеней свободы, которое имеет нервная система, за несколько сот миллисекунд редуцируется до одного и специфического состояния? И те, кто занимается когнитивными архитектурами мозга, думают, что решение этого вопроса лежит в том, что нервная сеть мозга организована не как случайная или не как регулярная сеть, а по принципу сетей малого мира, где локальная организация многих участков в одних небольших кластерах сопровождается длинными аксонодистантными связями с другими участками. Так что информация из одного участка за очень небольшое количество переходов (6-7 переходов) способна достичь любой другой точки нервной системы. Поэтому даже если нервная сеть оперирует очень медленно, 500-герцево, она имеет возможность глобального доступа. Для того чтобы воспроизвести это, помимо микроархитектуры с синаптической пластичностью и мемристорами нужна макроархитектура. То есть нужно организовать эти сети на основе сетей малого мира, и поэтому ведутся исследования, как можно это сделать.

Еще одно направление, которое нужно рассмотреть в связи с этим – это кортикоморфные когнитивные архитектуры. Другое военное ведомство США IARPA (Intelligence Advanced Research Projects Activity), ведомство, связанное со стратегическими исследованиями в области разведки, запустило другую программу, под названием MICrONS, которая расшифровывается как Machine Intelligence from Cortical Networks.

За этим стоят идеи, которые, кстати, обсуждались на этой конференции по когнитивному вычислению 10 лет назад в IBM. Там выступал Хоукинс, который развивает модели кортикоморфных вычислений. Идея заключается в следующем: кора головного мозга любого организма, обладающего ею (т.е. млекопитающего от крысы до человека), построена изоморфным образом. Несмотря на то, что количество нейронов у нас и у крысы сильно отличается, вся кора состоит из вертикально организованных модулей – колонок. Микроколонка содержит около 110 вертикально расположенных клеток, которые имеют стандартно построенную внутреннюю архитектуру, варьирующуюся немного от области коры к коре. Эти 110 нейронов есть в коре и у мыши, и у кошки, и у человека. Количество модулей увеличивается, то есть кора головного мозга человека располагает гораздо

большим количеством модулей и их специализацией, что позволяет ей выполнять очень многие задачи. Но идея кортикоморфных вычислений заключается в том, чтобы создать новое поколение машин, в которых алгоритмы обучения формирует обучающаяся структура. Такие машины будут достигать человеческого уровня компетенции за счет использования универсальных для любой коры всех млекопитающих примитивов как основной операции. И для этого исследуется, каковы могут быть эти примитивы, сколько находится нейронов в одной корковой колонке, как они связаны, каковы их входы и выходы, и что делается внутри каждой колонки с точки зрения логики переработки информации. И это через методы обратной инженерии переносится на создание таких вычислительных модулей, которые можно мультиплицировать, каждому давать свою вычислительную задачу (такие как, например, распознавание элемента паттерна или своего паттерна), а все вместе они могут выполнять сложнейшие операции, которые и обозначаются как когнитивный компьютер.

Мы тоже занимаемся этим. Год назад Министерство образования и науки начало проект «Разработка методов и платформ для исследования функционирования нервных систем на основе создания высокоразрешающей информационной модели кортикальных структур мозга». Мы выполняем там часть, которая преследует такие же цели. Мы задались вопросом, который принципиален для всех этих вещей, и в мире еще не решался. Ясно, что эти модули, если мы разделим кору на сотни таких модулей, каким-то образом перерабатывают и записывают информацию, а потом ее извлекают. Но так и не решен главный вопрос: в какие клетки (если мы все их знаем теперь, и они помечены специфическими маркерами) авансируется новая информация, то есть – куда попадает память, во все ли клетки? Любая ли клетка коры головного мозга может сохранить новую память или существуют только определенные клетки, способные на это? Если это определенные клетки, то какие именно их типы: возбуждающие, тормозные, длиннопроекционные, идущие в другие области и связывающие, или вставочные нейроны? Какие операции они выполняют во внутренней архитектуре? Мы в свое время разработали набор методов, позволяющих делать уникальные вещи. Они дают возможность визуализировать корковые модули и то, в какие клетки в них попадает новая информация. Мы имеем маркеры, которые позволяют прокрасить такие нейроны, которые запоминают новую информацию,

специальными флюоресцентными зондами и узнать, где они расположены, и как запоминание информации происходит.

Сочетая это вместе с набором специфических молекулярно-генетических и белковых маркеров типов нейронов, мы сейчас ведем исследования, которые призваны построить некоторые предварительные принципиальные схемы (нашего прошлого года работы): в какие нейроны записывается новая информация в коре головного мозга, что это за нейроны, где они расположены, есть ли специфика по слоям и ряд других вопросов.

Это текущий проект. Мы относимся к нему с очень большим интересом. И, я думаю, даже в проектах кортикоморфных вычислений, которые сейчас проводятся в Алленовском институте [головного мозга], этот вопрос не исследуется.

Нейронные гиперсети

Тем не менее, я хочу перейти к другому принципиальному вопросу. Несмотря на большие и впечатляющие достижения, которые были совершены в последние 10 лет в разных проектах, нет ответа на принципиальный вопрос о когнитивных вычислениях, результаты исследований говорят лишь о разных операциях в нейронных сетях коры головного мозга и в других структурах головного мозга. Откуда там берется когнитивность, откуда там берется разум и интеллект, эти разработки не объясняют.

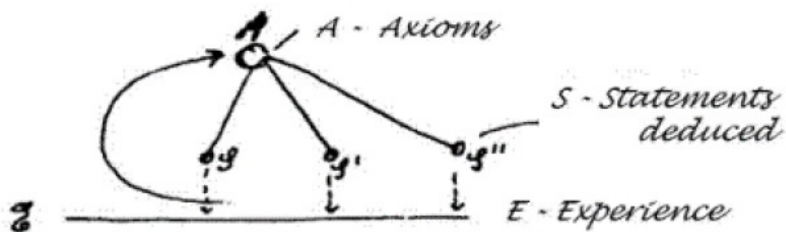
Моя установка в настоящем сообщении такая, что мы не решим проблему биологоподобного ИИ без нейронаучной теории, т.е. физикалистской теории естественного интеллекта, которая при этом должна быть не просто физикалистской; она должна решать фундаментальный вопрос о том, откуда из операций нервной системы берутся свойства разума и сознания. Любая теория, которая будет описывать операцию нейронной сети насколько угодно эффективно с точки зрения передачи сигналов, но которая не будет объяснять, откуда берутся когнитивные и субъективные свойства объективной реальности разума и сознания, не может быть эффективной. Нам нужна единая теория. Термин «I» в проблеме мозга и разума так же важен, как и в проблеме теории мозга или отдельной теории разума. А это философская проблема.

Но она же и научная проблема. И она требует определенной методологии. Я думаю, что сегодня мы в области нейронаук, с одной стороны, находимся в глубоком кризисе, замаскированном потрясающим взлетом исследований в этой области, но не отвечающим

на фундаментальные вопросы. А с другой стороны, стоим перед тем, что дальнейшее наращивание новых эмпирических фактов не решит этой проблемы.

Нам нужен поиск новых парадигм. А это требует новой методологии. Это требует методологии, которая должна привести нейронауку к таким же подходам, какие развивались и в физике. Иначе говоря, переход от простого сбора аналитического материала к масштабным теоретическим моделям. Сейчас опубликовано по отчету программы Human Brain Project около 30 млн. статей, имеющих отношение к проблеме мозга и разума. Но прибавление других сотен и тысяч не изменит ситуацию. Необходим другой модус работы, который ведет к абстракции, к идеализации, к формулировке аксиоматических принципов и выведению из них следствий, проверяемых в эксперименте. Эксперимент должен начать работать и в этом новом модусе.

Я привел для иллюстрации простую диаграмму (см. ниже), где несколькими росчерками пера Альберт Эйнштейн в 1952 году в рисунке своему другу юности Соловину описал то, каким образом строится научная теория. Нижняя черта – это эмпирический опыт, который мы имеем. Летящая линия собирает большое количество элементов из разных частей опыта, но не выводит какого-то из них индуктивно. Он назвал это «I», имея в виду, что это «*imagination*» или «*intuition*». Это не индуктивная логика, которая ведет к комплексу того, что он назвал. А это набор новых понятий, описывающих ранее скрытую реальность, и аксиом, из которых выводятся следствия. Часть из них находится в области фактов и объясняет существующие факты, а часть из них предсказывает факты еще не наблюдаемые, которые требуют новых экспериментов.



Четыре года назад я сделал задачу исследований в таком модусе предметом ежедневной работы. Она сложная, потому что должна

у любого экспериментатора чередоваться с экспериментальной работой, массой других дел, и это движущийся процесс. Я думаю, что формулировку проблем в этой области нужно перевести на такую же терминологию, как создание компьютерных программ: бета-версия, версия 1.0, 1.1 и так далее. Они меняются. Я бы считал, что то, что существует сейчас – это некая бета-версия этой теории. Тем не менее, она приобретает свои черты и облик. Очевидно, что такая теория, которая должна отвечать очень большому количеству требований, должна решать психофизиологическую проблему. Она должна апеллировать, с одной стороны, к феноменологическому материалу, объясняя факты феноменологии, идущие от первого лица, – вещь необычная для естественнонаучных дисциплин. С другой стороны, она должна апеллировать к фактам, идущим от третьего лица, которые касаются и психологии классической (со стороны третьего лица), и теории поведения, и фактов поведения когнитивных агентов, и мозга, объединяя все это.

В конечном счете, мне представляется, что такая теория должна проходить стадию трех моделей.

Первая – это принципиальная аксиоматическая модель, в которой такие принципы формулируются на еще не формализованном языке. Далее в работе с математиками она должна подвергаться формализации и превращаться в математическую модель. И дальше, что очень важно (хотя при математической модели возможны уже эксперименты на проверку и выведение новых следствий), чтобы вернуть ее назад к реальности, без которой ни одна теория не может существовать, необходимо опять взаимодействие нейроученых с математиками, потому что ее нужно наполнять дальше биологическими или физическими предсказаниями, переводя математические вещи в заземленные субстанцированные предсказания относительно механизмов, проверяемых в эксперименте.

Я условно или принципиально называю эту концептуальную новую модель теорией когнитивной. Лучший аппарат, который к настоящему моменту удалось найти для необходимой формализации – это теория гиперсетей. Я скажу о ней несколько слов. И, наконец, биологическая или субстанцированная модель – это гиперсетевая теория мозга.

Теория пользуется своего рода триадами, которые получаются естественным образом. Она ставит три принципиальных вопроса, которые должна решать любая такая теория: «что такое мозг?», «что

такое разум?» и «как соотносятся мозг и разум?». Она отвечает на них следующим образом. Любой мозг – это сеть. Это знали нейрочеловеки на протяжении последних 100-130 лет с открытием клеточной теории синапсов и нейронов. Следующий тезис менее тривиален и в когнитивной науке он не разработан, хотя существуют понятия о языковых сетях и семантических сетях, но это далеко до теоретического обобщения. Любой разум можно представить в виде сети, любой разум является сетью. И вопрос о соотношении разума и мозга решается в теории через гиперсетевую модель: разум – это гиперсеть над нейронной сетью, это сеть сетей из элементов нижней сети, это макроструктура. При этом мышление и сознание – это особый вид трафика, образующийся в макросети. Очень важно, что, если нервная сеть формируется в ходе развития мозга и готова почти полностью (с точки зрения количества нейронов, которые есть у новорожденного ребенка; после рождения появляется небольшое количество нейронных узлов сети в двух небольших структурах сети головного мозга), то когнитивная сеть начинается почти что с белого листа, хотя есть архетипические врожденные элементы этой сети, но она растет огромным образом. В эту сеть входят самые разные виды когнитивных элементов, о которых мы будем говорить, но в детстве они накапливаются потрясающими темпами. Ребенок учит в эти пики по одному новому слову и понятию в час.

Теория простая, как и должна быть любая теория в этой области, неважно, правильно ли это или нет, потому что нервная система возникла более 500 млн. лет назад и развилась в самых разных таксонах в разумное поведение. Она начиналась с очень простых нервных систем, и там уже должны были существовать базовые принципы и предпосылки для развития в самых разных ветвях эволюции свойств интеллекта и разума. Теория говорит, что для того, чтобы возник разумный или когнитивный агент, нужно всего три вещи. Во-первых, нужно, чтобы адаптивный агент состоял из функциональных систем – классических систем взаимодействующих элементов, которые приносят агенту адаптивный результат соотношения со средой. Кстати, понятия агента и среды являются критически необходимыми для эволюции этих систем.

Второе. В какой-то момент функциональные элементы должны заселить нервную сеть. Нервная сеть была великим изобретением эволюции. Она позволила делать очень многие вещи, в частности,

связывать разные когнитивные модули и элементы естественным образом.

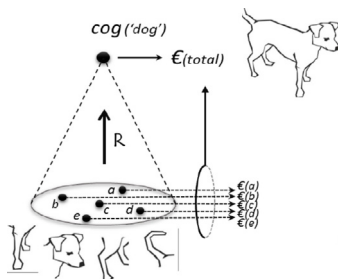
И третье условие: внутри нервной сети должны были возникнуть механизмы долговременной памяти, генетической пластичности, которая позволяет фиксировать процессы в нервной сети, переводя их в структурные состояния. Теория говорит, что разум будет эволюционировать и эволюционно возникать в любом адаптивном агенте, имеющем эти три условия. Нужно только время, проблемная среда и естественный отбор.

В теории существует три ключевых понятия. Теория вводит, как любая теория, сталкивающаяся с описанием реальности, требующая своего понятийного аппарата, три базовых понятия. Первое из них – это когнитивная группа. С одной стороны – это группа нервных клеток. С другой стороны, в совокупности она представляет собой элемент опыта, который кодирует тот или иной аспект соотношения когнитивного агента с теми или иными аспектами предметной среды. За счет того, что разум состоит из таких элементов (если хотите, монад), разум гранулярен, он дискретен. Но следующий пункт, за счет того, что это происходит в нервной сети, где элементы, образующие эти когнитивные группы, исходно связаны между собой, разум увязан. Он эти *коги* (cogs) образует между собой в устойчивые группы *links of cogs* – *локи*. И следующий важный момент: когда мы имеем такую систему, дифференцирующуюся внутри нервной сети, то она приобретает качественно новые свойства. Она имеет другую топологию, например, своих отношений, организации, доступов от одной группы к другой, и описание этой топологии это нечто большее, чем просто топология нервной системы. То есть так же, как у языка появляются новые системные свойства как у системы (а язык является одной из разновидностей или представителей этой сети), но точно так же и у всего разума как у когнитивной системы появляются новые сетевые свойства, которые могут быть исследованы, формализованы, просчитаны, измерены и описаны в рамках современных теорий сложных сетей и теории графов. Это расшифровывает понятие когнитивной группы.

Cog в английском языке – это элемент, часть шестеренки, а, с другой стороны, – это маленький элемент целостной системы. В терминологии теории когнитивной системы *cog* имеет два смысла. Первый – это некоторая вершина, образуемая большим количеством распределенных элементов, несущих общие новые свойства. А с другой стороны, аббревиатура *cog* – это когнитивная группа, это группа нейронов. Такое

решение не случайно. Оно адресовано к решению центральной психофизиологической проблемы, которую должна решать любая теория. Теория когнитума утверждает, что элементы гиперсети (*cogs* и *locs*) являются дуалами. Они, с одной стороны, являются физическими элементами нейронной сети, а с другой стороны, их совокупность имеет макросвойства, которые причинно лежат на более высоком уровне (и описываются, например, теорией интегрированной информации Дж. Тонони), чем набор самих физических элементов. И эта дуальность является принципиальной стороной и наших когнитивных связей в мозге, и наших когнитивных элементов опыта.

Если мы перейдем к необходимости формализации этого, то любой *cog* может быть описан как гиперсимплекс в гиперсети. Гиперсети – это вообще математический формализм, который не принят в обычной теории сетей. Он входит в область, даже скорее приходит из области топологической алгебры. Поэтому гиперсети состоят из симплициальных геометрических фигур, которые известны как гиперсимплексы. Потому что, с одной стороны, в основании лежит симплициальный элемент, симплициальный комплекс, а с другой стороны, вершина образуется через γ -реляционные отношения, эксплицитные описания этих отношений, которые имеют новые интегральные свойства, делающие ее элементом сети более высокого уровня.



Например, собака не расположена ни в одном или сумме просто элементов гиперсимплекса сети нижнего уровня, так же, как и три брикета, расположенные друг над другом, не образуют, как таковые, на нижнем уровне, понятие «ворота».

Если двигаться в сторону этого формализма дальше, то гиперсети представляют собой расширение более традиционных понятий графов и сетей, которые генерализуют их до гиперграфов и более

структурированно до гиперсетей. В отличие от графов и сетей, которые не только не способны описывать структуры более высокого уровня, у них нет реляционного аппарата формализации. Гиперсети способны описывать взаимодействие элементов не только попарно, что ограничивает графы в теории сетей, где две вершины связаны ребром.

Гиперсети – это сети сетей, и в теории когнитома (или теории нейронных гиперсетей) вводятся три клеточных правила, которые на системном уровне генерируют способность создавать все новые и новые узлы в гиперсети по мере приобретения нового опыта. Базисное из них для протокогнитивных групп – это пролиферация. Пролиферация – это когда существующая функциональная система, не имея способа достижения результата в репертуаре, который есть у когнитивного агента, ищет новый материал, являющийся средством и способом достижения этого результата. Но через некоторое время может автономизироваться и становится самостоятельной целью и задачей. Мы знаем массу таких примеров из области психологии.

Принципиально другой тип формирования когнитивных элементов относится к огромному количеству объектов психологии, начиная от концепций и кончая языком – это интергруппы, которые в теории описываются как квалоны, так как подразумевается, что они имеют некоторые качественные характеристики. Квалоны образуются в нейронных группах на местах перекрытия двух функциональных систем, когда одни и те же нейроны вовлекаются и в один вид деятельности, и в другой.

И, наконец, третий вид групп – это гипергруппы. Они возникают, когда много мелких и разветвленных когнитивных групп в разных областях мозга объединяются в целостные комплексы, имеющие свойства целостных сцен, целостного восприятия, целостного субъективного опыта. Теория говорит, что в любом когнитивном агенте есть два макроуровня – это альфа-когнитивные группы (классические функциональные системы) и бета-когнитивные группы (мелкие группы, которые могут иметь локализацию и когнитивную специализацию). И в некоторых когнитивных агентах (не во всех) эти микрокогнитивные группы (бета-модули) способны объединяться в крупные комплексы, дающие целостное восприятие, используемое когнитивным агентом для адаптации. Потому что сцена или целостный образ имеет свои качественно адаптивные характеристики по сравнению с набором признаков.

Я назову два важных свойства такой архитектуры, из этих свойств вытекает много следствий. Например, одним из принципиальных следствий является то, что формирование таких единиц в системе объясняет нефизическое субъективное время системы. Формирование каждого нового *cog*'а когнитивной группы, или каждого нового *loc*'а новой связи – это системное событие в жизни системы. И с помощью формирования этих событий измеряется собственное системное время, которое может отличаться от физического времени.

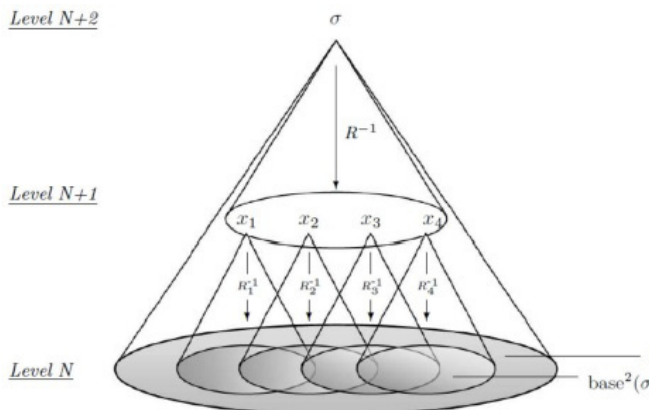


Fig. 9.10 $\text{base}^2 \sigma = \bigcup_{x_i \in R^{-1} \sigma} R_i^{-1} \sigma(x_i)$

На этой диаграмме я хотел бы подчеркнуть две вещи. Первое, что все эти вершины (классических функциональных систем, которые обеспечивают поведение) – микромодули, образующиеся на пересечении этих систем, и эти макровершины, гипермодули целостных состояний, которые включают эти микромодули, они все являются нейронными группами. Основанием всех этих групп являются нервные клетки, но свойства у них качественно разные. Причем то, что связано с субъективным опытом, не может возникнуть, пока не возникнут эти группы. И это не просто объединение групп. Они называются гипермодули, потому что это группа, состоящая уже не из нейронов, а из их групп, это группа групп. Последние содержат в себе как целое набор качественных квалонных признаков, которые могут отражать и сенсорные, и моторные, и многие другие компоненты.

Другая важная вещь, что за счет этой дробной процедуры, при которой возникают новые функциональные системы путем агрегации, затем путем возникновения новых функциональных систем, какие-то из элементов вычлняются и появляется дезагрегация, которая дает большое количество мелких групп. Надо понимать, что две функциональные системы, когда они работают на разных уровнях мозга, могут перекрыться в тысячи мест. И каждое из этих мест будет дифференцировкой соответствующих, лежащих в этой области нейронов, обладающих определенной специализацией. А затем новые агрегации – реагрегации – дают комбинаторный взрыв. И это помогает решить проблему, которую Хомский поставил перед когнитивной наукой как проблему дискретной бесконечности – каким образом небольшое количество стартовых нервных элементов способно дать бесконечное количество комбинаций. Теория говорит, что это происходит из-за того, что существует три процесса. Сначала происходит интеграция в крупные группы, потом существует дезинтеграция, или дезагрегация на мелкие группы, которые разбиваются и их становится огромное количество, а дальше из этих групп могут формироваться гипергруппы, которые комбинаторно дают невероятное количество комбинаций, обеспечивающих уникальность каждого момента нашего субъективного опыта.

И вторая вещь – вся эта система связана. Поскольку узлами являются нервные клетки, возбуждением отдельных из их связей мы можем достичь почти моментального проникновения, просачивания активности во все другие элементы этого когнитивного слоя (когнитивной гиперсети) за очень короткое количество переходов. Это своего рода кротовые дыры, туннели, которые существуют в нашем индивидуальном опыте через возбуждение этих общих нейронов.

Теория рождает много исследовательских вопросов, которые необходимо теперь переводить на уровень экспериментов. Мы исследуем это очень активно на уровне клеточных механизмов. У нас есть основания думать, что мы за счет предсказаний этой теории вскрыли совершенно новые принципы формирования ассоциаций между клетками, которые могут отличаться от классических павловских механизмов. Но это уже область экспериментальных исследований.

Заключить мне хочется тем, что я несколько раз был на лекциях Фрэнсиса Крика в начале 1990-ых годов, где он выступал с призывом исследовать сознание, что было совершенно нетривиально для нейрофизиологов того времени. И он всюду заканчивал одним

и тем же слайдом, на котором было написано «Сознание – сейчас». Этим он пытался подчеркнуть, что возможности для нейрофизиологических и нейронаучных исследований уже наступили. Вопрос в том, чтобы переломить психологический климат нейрофизиологов и перейти к исследованию сознания как научного вопроса. Это основное убеждение моего сегодняшнего доклада, что конструктивную теорию мозга и разума можно и нужно создавать сейчас. И имея ее в виде некоего прототипа, мы можем переходить к вопросам о том, как принципы логических тьюринговских машин или фон-неймановской архитектуры, заложенные 60-80 лет назад, могут быть заменены логикой когнитивных вычислений на материале реальной структуры нервной сети.

«УЛУЧШЕНИЕ» ЧЕЛОВЕКА И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ «IMPROVEMENT» OF MAN AND ARTIFICIAL INTELLIGENCE

幼丁 ««人类改善»与人工智能»

Юдин Борис Григорьевич

член-корреспондент РАН, доктор философских наук, профессор, вице-президент Российского национального комитета по биоэтике РАН, главный редактор журнала Президиума РАН «Человек», г. Москва

Yudin Boris G.

Corresponding Member of the Russian Academy of Sciences, Doctor of Philosophy, Professor, Vice-President of the Russian National Committee for Bioethics of the Russian Academy of Sciences, Chief Editor of the journal of the Presidium of the RAS «Man», Moscow



Аннотация. Искусственный интеллект рассматривается как социотехническая мнимость. Изучаются проблемы искусственных систем в связи с развитием биотехнологий. Обозначаются основные направления междисциплинарной экспертизы искусственных систем.

Ключевые слова: искусственный интеллект, искусственная система, биотехнология, экспертиза технологического проекта.

Annotation. Artificial intelligence is seen as a socio-technical imaginary. The problems of artificial systems in connection with the development of biotechnologies are being studied. The main directions of interdisciplinary expertise of artificial systems are indicated.

Keywords: artificial intelligence, artificial system, biotechnology, experiment of technological project.

В конференц-центре государственного парка Асиломар, в городе Пасифик Гроув штата Калифорнии, США, недавно проходила конференция, на которой рассматривались социально-этические проблемы, связанные с разработкой новой биомедициной технологии рекомбинантных молекул ДНК. Было ясно, что открывается масса заманчивых перспектив, но, с другой стороны, оказалось, что есть серьезные риски, сопряженные с этой технологией. Тогда в качестве таких рисков видели прежде всего опасность того, что, в первую очередь, с помощью этих рекомбинантных молекул ДНК можно будет создавать новые какие-то молекулы, которые, выйдя за стены лаборатории и распространяясь в окружающей среде, будут обладать неизвестными свойствами, на которые неизвестно как отреагирует окружающая среда. В общем, возможны всякие риски были. В этой ситуации исследователи предложили объявить на некоторое время мораторий на исследование в области рекомбинантных молекул ДНК, с тем чтобы попробовать собраться и прикинуть, какие возможны риски и что можно им противопоставить.

И вот, проводилась Асиломарская конференция, в ходе которой были разработаны меры предосторожности. Эксперименты с рекомбинантными молекулами были разбиты на четыре группы. В четвертую группу вошли наиболее рискованные, их было решено не проводить, другие стали проводиться с соответствующими предосторожностями. И это был уникальный эпизод, потому что тогда было неожиданно, что сами ученые заговорили о каких-то рисках, которые возникают, когда развитие технологий может коснуться окружающей среды и общества, человека. Это было сорок один год назад.

Теперь другая часть этого примера — это то, что было в прошлом году. В прошлом году было интересное событие, связанное в некотором смысле с продолжением той же линии рекомбинантных молекул ДНК. Это уже на новом уровне, с колоссальным развитием биоинформатики, ведь за прошедшие четыре десятилетия много чего нового появилось. В начале прошлого года (2015) в Китае в Гуанчжоу были проведены такие эксперименты, когда было использовано 86 оплодотворенных яйцеклеток человека, и все эти оплодотворенные яйцеклетки были носителями гена бета-талассемии, тяжелого наследственного заболевания. Цель эксперимента состояла в том, чтобы с помощью новой технологии, которая по-научному называется CRISPR/Cas9, вырезать этот больной ген (а это моногенный дефект) и

поставить на его место здоровый. И оказалось, что в четырех случаях из восьмидесяти шести это получилось успешно. Оказалось, что в руках у исследователей (хотя это, конечно, надо дорабатывать) новая технология, дешевая и простая в применении, которая позволяет эти генетические дефекты излечивать.

Казалось бы, полный вперед, надо добиваться дальнейших успехов на этом пути, но сразу же возникли проблемы, о которых тоже стали говорить исследователи и научные сообщества. И было решено, как и в том случае, собрать международную конференцию, которая в начале декабря прошлого года происходила уже не в Асиломаре, а в Вашингтоне — столице США. Она была организована Китайской Академией наук, Академией наук США (Национальным институтом здоровья) и Лондонским королевским обществом. И проблема опять же была связана с социальными и этическими последствиями и рисками применения новых технологий. Здесь эти риски носили несколько иную природу: не столько риски для окружающей среды, сколько связанные с тем, что была ранее поставлена некоторая граница — не естественнонаучная, а этического плана, которая оказалась не такой уж и неизбежной.

Речь идет о том, что делали китайские исследователи: они брали фактически эмбрион человека и экспериментировали с эмбрионом человека, а в существующей системе норм (хотя они не во всех странах такие жесткие, это чаще бывает на уровне так называемого «soft law» — мягкого права) такого рода эксперименты считаются допустимыми только при определенных условиях. Во-первых, нельзя создавать эмбрионы специально для исследовательских целей, то есть чтобы оплодотворение производить в пробирке, получать эмбрионы и затем проводить на них исследования. Во-вторых, существует проблема вмешательства в так называемую зародышевую линию. У каждого организма есть два типа клеток: соматические и зародышевые клетки. Если мы что-то делаем с соматическими клетками, то это остается на уровне того самого индивида, которым мы манипулируем, и никаких последствий для других индивидов это не несет. Если речь идет о зародышевых клетках, то те изменения, которые вносятся в зародышевые клетки, будут дальше передаваться потомству. Допустим, мы «отремонтировали» этот ген, который был дефектным и определял это заболевание, бета-талассемию. Но неизвестно, был ли этот ген за что-то еще ответственным; может быть, это просто еще не выяснили. И если это будет передаваться по наследству, то потом

могут оказаться какие-то совсем неприятные последствия. Есть заболевание — тоже генетическое — серповидноклеточная анемия, с которым в свое время на Кипре, скажем, стали бороться: проводили скрининговые программы, аборт подвергались те, у кого единственные зародыши были носителями дефекта. С другой стороны, оказалось, что ген серповидноклеточной анемии определяет устойчивость к малярии, поэтому на Кипре сначала это стали делать, а потом задумались, надо ли это делать, всякие протесты пошли.

Здесь идет речь о том, что имеется вмешательство в зародышевую линию, но китайские исследователи сказали, что они берут клетку с этим нездоровым геном и продолжают эксперимент до тех пор, пока клетка не начинает делиться. Вот после того, как дефект поправили, клетка начинает делиться, доходит до стадии восьми клеток, в этот момент эксперимент прекращают, чтобы дальше она не развивалась. Считается, что такого рода вещами можно заниматься до того момента, как 12 дней исполнится зародышу; это считается приемлемым.

Другая сторона этого развития связана с технологией. По-научному, как я сказал, она называется CRISPR, а в обиходе ее стали называть редактированием генов или редактированием генома. И вот сразу приходит на ум такая штука: редактирование генома может предприниматься не только для того, чтобы излечить и устранить какой-то дефект на уровне зародышевой клетки, но и для других целей — скажем, для улучшения тех детей, которые будут рождаться с использованием этой процедуры редактирования генома. И здесь открываются богатые возможности, вокруг этого идет сейчас много споров.

Ситуация была такая: когда собрались исследователи на международную конференции (назвали ее саммит, с первого по третье декабря), они приняли такое решение, которое не является обязательным (это опять же «мягкое право», с одной стороны, необязательное с юридической точки зрения, но с другой — нарушать его непристойно и неприлично), мол, надо поощрять развитие этих исследований, но при этом иметь в виду, что последствия не только в том, что 4 из 86 оказались удачны, но и что пошли мутации совсем в других генах. То есть эту технологию надо еще отрабатывать, она несет в себе такого рода риски. Ну и особая оговорка по поводу экспериментов с вмешательством в зародышевую линию: там довольно длинное рассуждение, смысл которого заключается в том, что пока на данной стадии

нельзя этим заниматься, необходимо проводить обсуждения, в том числе общественные, уточнения вводить, с тем чтобы когда-нибудь, возможно, получить добро на проведение таких экспериментов.

Я хочу обратить ваше внимание на то, что в этом случае, в прошлом году, это уже была совсем рутинная штука. Что я имею в виду: исследователи собрались и начали обсуждать социально-этические проблемы еще до того, как технология, что называется, пошла в массы. За 40 лет с 1975 по 2015 год произошел такой существенный сдвиг, что это стало нормой. Вот смотрите: когда был проект «Геном человека», который готовился в конце восьмидесятых-начале девяностых, тогда было решено, что этот проект будет финансироваться (основное финансирование было со стороны Штатов), финансироваться щедро – тогда это был самый богатый проект в истории мировой науки – с тем, что примерно от 3 до 5 % средств, которые идут на этот проект, должны идти на изучение юридических и социально-этических проблем, которые назывались ELSI (Ethical, Legal and Social Issues), связанные с этим проектом. То есть речь шла о том (и Асиломарский мораторий был первым таким случаем), что к рискам, связанным с новой технологией, надо обращаться не тогда, когда люди станут сталкиваться с ними в повседневной жизни, а когда технология эта только еще разрабатывается. Таким образом, за прошедшие десятилетия это стало нормой; механизм обсуждения проблем (ELSI в Америке) для того, чтобы общество, политики и сами научные сообщества были готовы и могли быстро реагировать в случае чего. Потом, в последние лет десять, те исследования, которые финансируются европейским сообществом, – это уже не США, это Европа – должны соответствовать требованиям RRI (ответственные исследования и разработки). Имеется в виду, что при рассмотрении какого-либо проекта проведения исследования или разработки новой технологии, уже на этой стадии необходимо заниматься – и это обязательное условие – изучением возможных социально-этических и иных последствий.

Таким образом, можно говорить о том, что в современной научно-технической практике формируется такой слой знаний, поскольку этим же надо кому-то заниматься – изучением этических и правовых проблем. Это становится отдельной сферой в рамках научной, технологической, проектной деятельности. Таким образом, образуется какой-то слой социально-гуманитарных знаний, который оказывается необходимой стороной сегодняшних исследований, проводимых

в самых разных сферах, но особенно в той мере, в какой они направлены на человека, то есть где человек является мишенью этих исследований. Далее, ситуация такая: я говорил, что все больше и больше технологий сосредотачиваются на человеке, и здесь исходная ситуация была такая, что биомедицинскими технологиями занимались давно, не то чтобы это дело каких-то последних десятилетий. А ведь они, по определению, — технология воздействия на человека. Воздействие на человека, имеющее целью устранение каких-то дефектов, поправку здоровья и так далее. То есть имеется в виду, что такие исследования и разработки направлены на получение терапевтического эффекта. Терапевтического не в том смысле, что есть врач-терапевт, а в смысле улучшения здоровья.

И это так и шло, были эти исследования, а в последнее десятилетие стало все больше разговоров (и не только разговоров) о том, что эти технологии, оказывается, бывают направлены не только на получение терапевтического эффекта, но и на «улучшение» человека. Когда программа составлялась, я подавал название своего доклада не в письменном виде, а просто на слух записали. И слово «улучшение» оказалось без кавычек, а надо было в кавычках. Здесь оказалось без кавычек, это какой-то мой дефект, но кавычки для меня важны, чтобы показать, что «улучшение» — вещь далеко не всегда прямая и односмысленная, и когда говорят об этом «улучшении», надо внимательно смотреть, является ли это действительно улучшением. Может, это для кого-то улучшение, а для кого-то нет. Даже вот эти побочные, вторичные эффекты, которые не предусмотрены вначале, какое они могут воздействие оказать?

И сейчас говорят об улучшении разного типа: улучшении физических кондиций человека, позволяющем сделать его более сильным физически, ростом повыше, и так далее. То, что сейчас широко обсуждается на самых высоких уровнях, что связано с допингом, — это тоже разновидность такого улучшения. Далее, эмоциональное улучшение (enhancement по-английски), чтобы человек легче в эмоциональном плане переносил разные стрессы, формирование психической устойчивости. И когнитивные улучшения; об этом особенно много сейчас говорят. И здесь есть какие-то реальные технологии, которые достаточно широко используются для такого рода улучшений.

Здесь возникает проблема, которая давно обсуждается: как провести границу между терапией, с одной стороны, и улучшением — с другой? Смотрите, когда есть терапия, и когда мы говорим о терапии,

мы вполне готовы согласиться с тем, что будет какой-то риск, процент неудачи, в общем, если речь идет о том, чтобы поправить здоровье, то мы на это будем соглашаться. А если это улучшение, когда я и так хорош, а мне хочется стать еще лучше, то здесь отношение к риску будет совсем другое. И это сказывается не только на оценке риска, но и на том, что, допустим, в странах, где государство оплачивает или есть медицинское страхование, покрывающее риски от терапевтического вмешательства, почему народ должен солидарно оплачивать ваше желание себя улучшить? Поэтому изучение этого вопроса о различении имеет практический смысл.

Провести границу, повторюсь, бывает очень и очень непросто. Вот такой пример, который часто рассматривается в литературе в связи с этим: есть такой препарат — риталин. Он был придуман и вошел в практику для борьбы с синдромом дефицита внимания. Синдром дефицита внимания бывает у мальчиков, которые очень непоседливы, им тяжело сосредотачиваться, и так далее. И от этого тяжело окружающим, тем, кто с ними в одном классе, учителям, родителям — в общем, все были согласны на то, чтобы что-то такое сделали. Считалось, что дети бывают разными и к некоторым надо просто более строго относиться, чтобы те исправились. Потом было установлено, что это есть синдром — синдром гиперактивности и дефицита внимания. Синдром — это уже совсем другое отношение, это значит, их надо лечить. И вот был создан риталин (сейчас, конечно, уже большая группа этих препаратов), для того, чтобы преодолеть этот синдром гиперактивности. А потом оказалось, что его стали применять самочинно студенты, которые считали, что таким образом это позволит им при подготовке к экзаменам быть более сосредоточенными и не отвлекаться на что-то постороннее, так как риталин позволяет добиться этого эффекта. Сейчас я не буду вдаваться в подробности о том, насколько это позволяет, насколько не позволяет, и какие есть другие пути, но это различие для меня важно: в одном случае имеется в виду терапевтическое воздействие риталина, а в другом — воздействие улучшающее.

Я хочу сказать следующее: создание новых технологий сегодня — это деятельность многоплановая, не так, чтобы в одной лаборатории придумали технологию, и дальше она пошла в жизнь. Это деятельность многоплановая, которая носит в большой мере не только исследовательский и разработческий характер, но и социальный, поскольку в этот процесс создания, применения и доведения до

потребителей этих технологий включены и бизнес, и лаборатории, и ученые-создатели технологий, и потребители, которые ждут появления новых технологий. Это очень интересная особенность сегодня: все больше технологий разрабатывается не потому, что пришла в голову какая-то идея, а потому что эта идея как бы носится в воздухе, и более того, она вполне оформлена. Есть какие-то движения потребителей, допустим, движение пациентов или родственников пациентов с болезнью Альцгеймера. Они, во-первых, выступают как лоббисты, во-вторых, зачастую и спонсируют проведение этих исследований по разработке и клиническим испытаниям препаратов, которые, как считается, позволят побороть болезнь Альцгеймера.

Вот это только один пример того, что технологии разрабатываются, когда уже есть соответствующие ожидания. И мне представляется, что если взять сферу искусственного интеллекта, то есть понятие, которое широко используется в современной социологии науки, то к нему применимо и понятие социотехнических мнимостей (sociotechnical imaginaries). Мнимости — то, чего еще нет, но тем не менее то, что реально участвует и во многом определяет не только действия отдельных индивидов, но и направления исследований. Мне представляется, что «искусственный интеллект» — это имя для определенной социотехнической мнимости.

Когда я употребляю это слово, я не имею в виду ничего предосудительного. Можно перевести это как «воображаемое», но, когда я математику изучал, там говорилось, например, о мнимых числах. Я к чему это говорю: многое в развитии современной науки и технологии, в том, что называется «технонаукой», определяется социотехническими мнимостями, которые, с одной стороны, что-то такое выдуманное, но с другой стороны, поскольку много людей этим занимаются, это приобретает реальные черты.

ФУЭТЕ, ФОНЕМА, ФОРМУЛА, ФОТОН: ЯЗЫКИ МОЗГА И КУЛЬТУРЫ FUETE, PHONEME, FORMULA, PHOTON: LANGUAGES OF THE BRAIN AND OF CULTURE

车尔泥古扶斯卡亚, 扶矮特 «音位, 公
式, 光子: 脑子和文化语言»

**Черниговская Татьяна
Владимировна**

*доктор биологических наук, доктор
филологических наук, заведующая лабораторией
когнитивных исследований и кафедрой проблем
конвергенции естественных и гуманитарных
наук Санкт-Петербургского государственного
университета, г. Санкт-Петербург*

Chernigovskaya Tatiana V.

*Doctor of Biological Sciences, Doctor of Philology,
Head of the Laboratory of Cognitive Studies and
the Department of Convergence of Natural and
Human Sciences of St. Petersburg State University,
St. Petersburg*



Аннотация. Обозначены междисциплинарные проблемы искусственного интеллекта на стыке психологии, лингвистики, биологии. Приводятся факты сравнительного изучения мозга приматов, ученых, танцоров, музыкантов. Поднимаются вопросы соотношения человеческой и компьютерной памяти.

Ключевые слова: когнитивная наука, психология, лингвистика, биология, нейронаука, мозг, память.

Annotation. Interdisciplinary problems of artificial intelligence are identified at the junction of psychology, linguistics, and biology. The facts of comparative study of the brain of primates, scientists, dancers, musicians are given. Questions of the relationship between human and computer memory are raised.

Keywords: cognitive science, psychology, linguistics, biology, neuroscience, brain, memory.

Клод Лелуш, известный французский режиссер, сказал следующее: есть только один гениальный постановщик, и это Господь. В его пьесе играют миллиарды людей, каждый уверен, что он в главной роли, и самое интересное, что каждый из этих людей прав, при условии, что этих людей миллиарды. Более того, нет режиссера настоящего как такового, пьеса запутанна, никто не знает сценарий, никто никогда не репетирует, все импровизируют и пытаются понять как свою роль, так и замысел пьесы в целом. На мой взгляд, это относится к социуму в целом и к мозгу в частности.

Я хочу напомнить, что Шредингер написал в 1944 году книгу «Что такое жизнь?», где он отвечает на этот вопрос с точки зрения физики. Он говорит о том, что мы пришли к такой точке, когда количество знаний таково, что мы не знаем, что с ними делать. Только я бы внесла тут поправку: количество не знаний, а данных. На мой взгляд, это очень важное замечание. От того, что суперкомпьютер совершает триллионы операций в секунду, знаний не добавляется. Знания появляются в результате анализа этих данных людьми. Моментальный поиск информации в интернете не увеличивает наш умственный уровень, а, возможно, даже и уменьшает.

Таковыми проблемами занимается когнитивная наука, о чем подробно нет времени говорить. Замечу лишь то, что она действительно междисциплинарна. Моя основная претензия к этой области знаний заключается в том, что почти не осталось дисциплин, которые с ней не связаны. Казалось бы, можно сказать, что металлургия к ней не относится, но это тоже будет неправильно, потому что можно придумать ход, когда будет описываться, как происходит процесс изготовления способа получения нового металла.

Пьер Тейяр де Шарден говорил, что эволюция жизни на Земле, насколько мы вообще можем о ней что-то знать и говорить, имеет такой вид – непрерывное усложнение систем, которое дошло до гиперсетей. В основе всего этого лежит способность человека к оперированию знаковыми системами. То есть мы, люди, живем, и, возможно, только люди, как минимум в двух мирах: мире материальном и мире знаковом. На самом деле, второй, скорее всего, уже играет для нас большую роль.

Почему мы так страдаем и переживаем, читая истории про братьев Карамазовых или про страдания юного Вертера, ведь их же никогда не было? Для мировой культуры и для каждого из нас

отдельно, если мы образованны, мир знаков играет огромную роль. Каждое живое существо, в особенности, разумные существа, отличается характеристиками мозга: объемом мозга и способом его организации. Когда разговор идет о системах искусственного интеллекта, которые могли бы имитировать интеллект, то речь все же идет о более низком, простом способе организации, нежели у человека. Между тем, сегодня ректор Московского университета говорил о шахматных программах, а неделю тому назад я говорила своим студентам, что есть такие вещи, через которые компьютер не в состоянии перешагнуть, например, игра в Го. Сейчас мы видим, что нейросеть Google обыграла чемпиона девятого дана по Го. На мой взгляд, я могу ошибаться, что дело тут уже не в количестве операций, которые просчитывает компьютер, а в виде ходов и принципах мышления или его имитации. В качестве контраргумента можно привести игру в нарды, где есть элемент случайности в виде бросания костей, и где я могу сделать «истерический ход», когда я хожу наугад и все равно выигрываю. Тут можно, конечно, сказать, что я все равно просчитываю ход, но делаю это неосознанно. Это — детали. Я лишь хочу сказать, что еще один рубеж в развитии искусственного интеллекта преодолен.

Есть еще одна интересная тема, которая до сих пор всех интересует. Как так получилось, что человек сделал такой большой рывок в своем развитии? Есть много спекуляций на эту тему, и вот одна из них в генетике. Отличия необходимо искать не в количестве совпадающих генов у человека и приматов, а как раз в том минимальном различии, которое мы имеем. Также есть мнение о том, что это зависит не от генов, а от неравномерного развития различных участков мозга. В этой же сфере исследователи делают упор на тот участок мозга, который отвечает за речь. Так как явно произошло какое-то отклонение, когда возник человеческий язык. Обращаю ваше внимание на то, что человеческий язык именно возник. Но не в том смысле, в котором говорит Хомский, говоря о макромутациях. Тогда что значит «возник», когда у всех животных есть языки?

Когда я говорю «язык», я имею в виду только человеческий язык. Хотя такая предпосылка сейчас тоже может считаться ошибочной, потому что три дня назад оказалось, что нашли синтаксис у некоторых видов животных. У нас в голове есть база универсальных правил, которая едина для всех языков планеты, которых 6-7 тысяч. Несмотря на огромную разницу между языками, есть базовые правила для

языка *homo sapiens*. Похоже, что эти базисные правила врождены и даже генетически обусловлены. Ребенок рождается в языковой среде, о которой он не знает ничего, и получается, что если он не имеет этих базисных правил, то он совсем не способен научиться языку, а если же они уже есть, то этого недостаточно, и ему необходимо получить «прививку» языка, в среде которого он родился. Его мозг вынужден расшифровывать этот код, с которым не может справиться вся лингвистика, включая Хомского. Мозг ребенка делает это в течение двух-трех лет, если он здоров. Он как бы пишет себе виртуальный учебник данного языка и пользуется им и начинает таким образом говорить. Когда эти базисы для всех языков зарождались, происходили изменения не анатомические, а функциональные, когда Создатель или эволюция решили редуцировать большие функции мозга, то есть разделить их на куски. Только не стоит считать, что эти куски обозначают физические модули. Томография и магнитно-резонансные исследования помогают увидеть нам сложнейшую сеть, которая задействована всегда, и по ней невозможно определить, где кончается внимание, где память, а где язык.

Соссюр, Jakobson и Хомский говорят о языковой типологии: что в языках общее, а что различно, о языковых семьях и т. д. Когда же генетика начала усиленно развиваться, то это помогло нам установить связи типов языков с этническими и генетическими особенностями. Если мы будем говорить про речь, то язык задействует различные области мозга. Основные дебаты происходят вокруг соотношения областей языка: морфологии, синтаксиса и т. д., с областями мозга, а также типах их организации: по модулям или сетевая.

Также ведутся споры между психологами и лингвистами. Связаны ли между собой психологические и лингвистические проявления сознания у человека? Например, те, кто говорят, что язык существует отдельно от психологических явлений, приводят в качестве аргументов генетические изменения и ошибки. В семьях, где у каждого из членов семьи все в порядке с интеллектуальным развитием, но у них не в порядке языковое развитие. И это также говорит в пользу модульного устройства мозга человека. У художников, творческих личностей иначе организован мозг.

В качестве примера хочу привести артистов балета. Ученых заинтересовала работа мозга танцовщиков, делающих фуэте и другие сложные и быстрые фигуры. Благодаря огромному количеству тренировок, артисты балета выполняют эти движения не задумываясь, не

прогоняя каждый раз в голове всю последовательность необходимых для этого движений. Но это не означает, что мозг танцора в этот момент не активен. Задействованы почти все области мозга, которые координируют каждое движения танцора. Музыка активизирует связи, идущие по мозолистому телу, связывающему левое и правое полушария. Математика и музыка – два мощнейших инструмента для тренировки нейронных связей в мозге. Математика тоже задает некую базовую структуру нейронов, а также наоборот: в мозге заложена способность к математике.

Исследования животных показывают, что некоторые виды обладают способностями к математике. Самое интересное, что не только высшие виды вроде приматов, но даже муравьи обладают такой способностью. Эксперименты с муравьями проводятся Ж. И. Резниковой в Новосибирском институте. При сканировании же мозга человека, например, сильнейших математиков, реакции мозга похожи на реакции на живопись, так как красота формул для них сравнима с восприятием живописи. Это зависит от способа и уровня восприятия математических объектов. При сканировании мозга математиков выяснилось, что плотность серого вещества у математиков больше, чем у обыкновенных людей. Главный вопрос тут: врожденно ли это или приобретено?

К примеру, при изучении мозга Эйнштейна выяснили, что мозолистое тело у него было значительно больше, чем у людей мужского пола его возраста. Многие считают, что причина тому, что Эйнштейн занимался музыкой, а именно, играл на скрипке. Здесь задают еще один вопрос: влияет ли вид инструмента на развитие нейронных сетей мозга?

Существуют исследования, которые показывают, что у музыкантов, играющих на клавишных, нет сильной асимметрии мозга, так как задействованы обе руки в равной степени. У музыкантов, играющих на струнных инструментах, видна асимметрия из-за особенностей моторной деятельности, завязанной на инструменте. У Эйнштейна был очень высокий индекс соотношений нейронов и глии. Извилины в мозгу, связанные с серым веществом, были необычные. Суть исследований мозга Эйнштейна заключается в фразе «неудивительно, что Эйнштейн был гением».

Вопросы, которые меня интересуют. Зачем человеку необходима такая большая нейронная сеть? Существует ли запас возможностей мозга? Однородна ли эта сеть? Когнитивное внутри клетки?

Сознание – эпифеномен или нет? Откуда взялась субъективная реальность? А есть ли решение психофизической проблемы? (я считаю, что нет, но Константин Владимирович Анохин вроде бы решил эту проблему) Структура равна функции? Мозг в последние годы называют биофотонной графической матрицей, квантовым компьютером или сетью. Насколько связаны визуальные картины в мозге со зрительными каналами?

Человек уже давно смастерил себе внешнюю память, которая находится отдельно от нас и не связана с мозгом. Сейчас же мы уже плохо понимаем, как работает внешняя память (облачные хранилища данных, big data и пр.). Уничтожение этой внешней памяти, также как и с мозгом, возможно при тотальном блэкауте, то есть путем лишения электрического снабжения мозга или машины. Информация пропадет.

В 21 веке человечеству пророчат переход в виртуальную реальность, которая станет приоритетом для человечества. Хотелось бы задать вопрос. Искусственный интеллект, который будет обгонять нас в развитии, якобы должен будет иметь собственное «Я»; сможет ли он страдать или умирать?

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ РОБОТОВ ПАРАЛЛЕЛЬНОЙ СТРУКТУРЫ В МЕДИЦИНЕ

THE POSSIBILITIES OF USING PARALLEL ROBOTS IN MEDICINE

各拉祖诺扶 «医学中使用平行结构机器人的能力»

Глазунов Виктор Аркадьевич

член-корреспондент РАН, доктор философских наук, доктор технических наук, профессор, директор Института машиноведения РАН имени А. А. Благонравова, г. Москва

Glazunov Victor A.

Corresponding Member of the Russian Academy of Sciences, Doctor of Philosophy, Doctor of Technical Sciences, Professor, Director of the A. A. Blagonravov Institute of Mechanical Engineering of RAS, Moscow



Аннотация. Обозначены междисциплинарные проблемы медицинской робототехники. Сравниваются особенности последовательных и параллельных роботов. Демонстрируются прототипы параллельных роботов.

Ключевые слова: медицина, робототехника, последовательная структура, параллельная структура.

Annotation. The interdisciplinary problems of medical robotics are indicated. The features of sequential and parallel robots are compared. The prototypes of parallel robots are demonstrated.

Keywords: medicine, robotics, sequential structure, parallel structure.

Я хотел бы представить некоторые работы нашего института машиноведения РАН имени А. А. Благонравова. Мы занимаемся, в том числе, робототехникой, хотя направления наших исследований весьма обширны. Должен сказать, что наш институт один из первых в мире начал исследовать роботов. В связи с этим я должен вспомнить работы сотрудников нашего института академика АН СССР Ивана Ивановича Артоболевского и профессора Арона Ефимовича Кобринского. Они впервые сделали протез руки, который работал от биотоков. Затем Арон Ефимович впервые в мире предложил станок с ЧПУ. Это был конец 1950-ых – начало 1960-ых годов.

Сегодня я хотел бы обратить внимание на работы в области хирургических роботов. Должен сразу сказать, что существует мнение, что это не роботы как таковые, и что это робото-ассистированные операции. Это манипулятор, которым управляет человек, хирург. Он может находиться как в нескольких метрах от робота, так и в тысячах километров, в зависимости от линий связи.

Возможно, аудитории будет интересно, что робот Да Винчи (рис.1), самый популярный робот американского производства, и, пожалуй, единственно работающий робот-хирург, разрабатывался достаточно долго, более 30 лет.

Эти роботом планировалось оснастить отдаленные базы и подводные лодки, и даже космические корабли, где можно было дистанционным образом оперировать пациентов. Однако в результате он получился довольно громоздким. Робот имеет четыре руки, каждая из которых весит примерно 100 кг.

На рис.2 главный уролог России профессор Дмитрий Юрьевич Пушкарь сидит за пультом и управляет роботом «Да Винчи».

На конце каждой из рук находится определенный хирургический инструмент. При введении в тело пациента необходимо этот инструмент четко позиционировать. Поэтому должна быть некоторая точка ввода, которая образуется за счет небольшого прокола в теле. Далее эта точка никуда не должна перемещаться. В этом «проколе» и состоит специфика работы данного механизма. При этом получается так, что рука, весящая 100 кг, перемещает груз, весящий всего лишь 100 гр и даже меньше. Соотношение один к тысяче. За счет чего это произошло?



Рис.1. Робот-хирург «Да Винчи»



Рис.2 Управление роботом «Да Винчи»

В нашем институте мы подробно исследовали эту структуру. И оказалось, что эта структура строго последовательна. Ее строение таково, каково строение руки человека. К тому же, из-за необходимости наличия фиксированной точки ввода, механизм должен иметь некоторую избыточность. Чтобы управлять системой по 6 степеням свободы, механизму необходимы 8 двигателей. Получилась негативная энергетическая зависимость веса инструмента от поставленных задач.

Я называю это устройство «роботом», несмотря на то, что он, скажем прямо, не очень-то и думает. Почему это все-таки робот, хотя он и непосредственно управляется человеком? Дело в том, что роботу приходится «понять», что хочет хирург-человек и перевести «желания» хирурга, выраженные в человеческих движениях, в движения другого рода. То есть, ему надо пересчитывать и вычислять необходимые характеристики собственного движения. Вычислить, как требуется переместить инструмент в соответствии с движениями хирурга и, далее, рассчитать работу электромеханического привода, который осуществляет движения хирургических инструментов в соответствии с рассчитанными траекториями. Это так называемая прямая задача кинематики. То есть, в принципе, хотя и налицо непосредственное взаимодействие человека и машины, однако все равно присутствуют роботизированные схемы движений.

Мы сразу стали исследовать альтернативные варианты построения роботов. Имеются проекты роботов т.н. параллельной структуры.

В отличие от последовательной структуры, здесь имеется несколько кинематических цепей, соединяющих рабочий орган с основанием. При этом посредством параллельного пересечения нескольких цепей получается довольно-таки жесткая конструкция, которая необходима для точного направления хирургических инструментов.

На рис. 3 представлен параллельный робот для осуществления стоматологических операций. На рис. 4 — робот параллельной структуры польского аналога робота «Да Винчи». В ней фиксация инструмента задается не за счет опорной точки и дополнительных приводов, как это было у американцев, а за счет синтезирования конкретной структурной схемы эффективного управления.

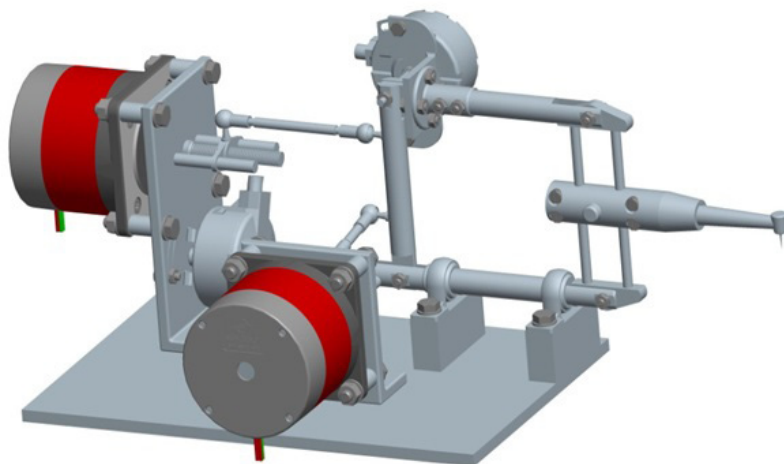


Рис.3. Устройство параллельной структуры для ортопедических операций

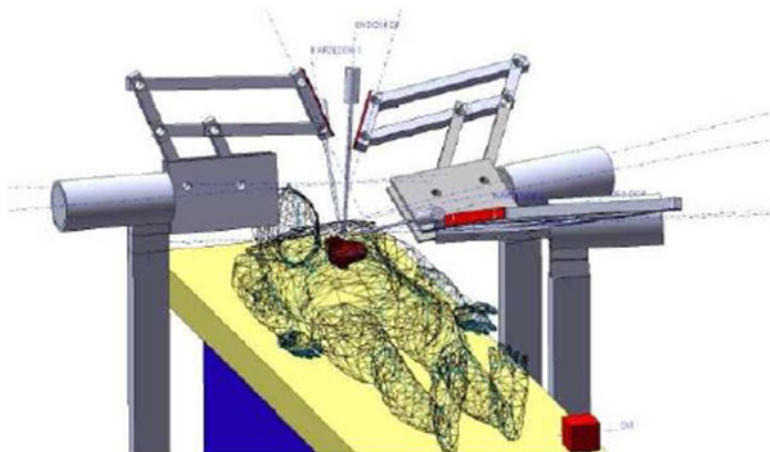


Рис.4. Робот параллельной структуры – альтернатива роботу Da Vinci (Польша)

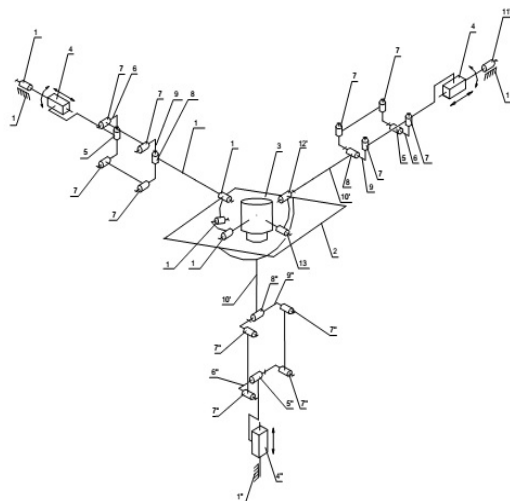


Рис.5. Механизм с дополнительными приводами робота «Ортогляйд» (5 степеней свободы)

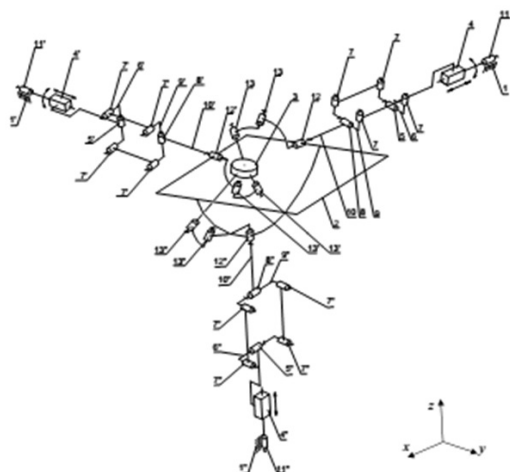


Рис.6. Механизм с дополнительными приводами робота «Ортогляйд» (6 степеней свободы)

На такой идеологии строятся и наши работы. Суть ее в том, что для каждой специфической работы робота, под каждую конкретную задачу нужно подобрать точную оптимальную структурную схему параллельного управления для задач динамики. Это осуществляется на самых ранних стадиях проектирования робота.

В нашем институте решаются различные задачи. Многие проекты выполняются совместно с зарубежными коллегами. Так, в Институте кибернетики и коммуникаций города Нант во Францию был построен робот «Ортогляйд» («ORTHOGLIDE»). Этот робот производил чисто поступательные движения, без смены ориентаций (рис. 5). Нам удалось, в рамках совместного проекта, за счет дополнительных двигателей и путем модификации структурной схемы, добавить роботу свободу вращения (рис. 6).

Все вышесказанное, может, и не совсем относится к искусственному интеллекту. Тем не менее, алгоритмы, которые мы применяем в робототехнике, на мой взгляд, подчеркивают высокий интеллект разработчиков. Роботы работают параллельно. Человек не способен осуществлять параллельные действия.

ПАРАДИГМАЛЬНЫЙ АНАЛИЗ ТЕОРИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

PARADIGM ANALYSIS OF THE THEORY OF ARTIFICIAL INTELLIGENCE

库姿妥册扶 «人工知能理论的范式分析»

Кузнецов

Валерий Григорьевич

доктор философских наук, профессор,
заведующий кафедрой философии и методологии
науки философского факультета МГУ имени
М. В. Ломоносова, г. Москва

Kuznetsov Valery

Doctor of Philosophy, Professor, Head of the
Department of Philosophy and Methodology of Science,
Faculty of Philosophy, Lomonosov Moscow State
University, head of the seminar «Neurophilosophy»,
Moscow



Аннотация. В статье описывается история становления и развития символического и коннекционистского подходов в теории и технических приложениях искусственного интеллекта. Методологической основой работы является парадигмальная философия науки Томаса Куна.

Ключевые слова: символизм, коннекционизм, парадигма, Кун.

Annotation. The article describes the history of the formation and development of symbolic and connectionist approaches in the theory and technical applications of artificial intelligence. The methodological basis of the work is the paradigmatic philosophy of science by Thomas Kuhn.

Keywords: symbolism, connectionism, paradigm, Kuhn.

История возникновения когнитивизма и искусственного интеллекта как взаимозависимых направлений имеет глубокие корни. Если проследить истоки, формирование предпосылок когнитивных наук и первых программ искусственного интеллекта, то история философии и история науки (прежде всего биологии, психологии и физиологии) даст нам большое количество сведений, освоение которых является самостоятельной задачей. В данной статье я хотел бы рассмотреть конкретную проблему возникновения когнитивистской гипотезы, как она формулируется, как связана с искусственным интеллектом и философией. Главной интуицией когнитивистской гипотезы является положение, что мышление есть вычисление, а человеческий разум – устройство для получения, хранения, переработки и использования информации. Основания для этой гипотезы были заложены Г. Лейбницем, Т. Гоббсом, Д. Булем, А. Тьюрингом, К. Шенноном¹. Когнитивистская гипотеза есть попытка связать познавательную деятельность человека (и не только человека) с работой головного мозга как вычислительного устройства. Фактическая связь между когнитивистской гипотезой и методологией искусственного интеллекта в западной науке и философии была реально обозначена в 1956 году на двух конференциях в Кембридже и Дартмуте. Спустя 35 лет Ф. Варела, Е. Томпсон и Е. Рош в своей книге «Воплощенный разум: когнитивная наука и человеческий опыт» так описывают это событие: «Как было очевидно ясно, что кибернетика возникла в 1943 году, так же было ясно, что когнитивизм возник в 1956 году. В течение этого года, на двух конференциях, состоявшихся в Кембридже и Дартмуте, новые голоса (такие, как Герберт Саймон, Ноам Хомский, Марвин Минский и Джон Маккарти) выдвинули идеи, которые должны были стать основными принципами для современной когнитивной науки.

Центральной интуицией когнитивизма является то, что интеллект и включенный в него человеческий интеллект столь напоминает вычисление в его сущностной характеристике, что познание действительности может быть определено как вычисление символических представлений. Очевидно, что это направление не могло появиться без основания, установленного в течение предыдущего десятилетия.

Главное отличие в том, что одна из многих оригинальных предварительных идей теперь была повышена до полноценной гипотезы,

¹ См. Меркулов И. П. Когнитивная наука // Новая философская энциклопедия. Т. 2, М., 2001, с. 164-165.

с сильным желанием расширить ее границы, ее приблизительные и междисциплинарные корни, где социальные и биологические науки получили бы преимущество при всей своей многогранной сложности»¹. Так, в частности, головной мозг человека, сформировавшийся в ходе биологической и культурной эволюции, может быть интерпретирован как вычислительное устройство по аналогии с операциями, осуществляемыми компьютером. Отсюда вытекает первая версия понимания искусственного интеллекта: компьютерные программы способны моделировать человеческие когнитивные способности. Такую версию, основанную на отождествлении функций человеческого мышления с работой центрального процессора в компьютерных системах, называли слабой. Она практически без возражений была принята среди представителей нейронауки и философов. Сильная версия допускает создание и существование вычислительных устройств, способных обладать разумом и пониманием². Но со стороны нейронауки даже слабая версия была подвергнута критике, т.к. она не соответствует данным нейробиологии. Критиками выступили сторонники коннекционизма, которые полагали, что познавательные способности живых организмов не могут моделироваться вычислительными устройствами на базе центрального процессора, они должны имитироваться, «симулироваться» с помощью искусственных нейронных сетей³. Такой подход подсказал, в частности, направление развития компьютеров нового поколения и вместе с этим новую стратегию развития методологии искусственного интеллекта.

Влияние когнитивизма на философию науки вообще и на определенные, связанные с искусственным интеллектом аспекты философии сознания в частности, на становление и развитие конкретных когнитивных нейронаук и теории искусственного интеллекта было весьма весомым и заметным. Развитие нейронауки после краха биохевиоризма (отрицавшего роль внутренних психических процессов) сначала было отмечено достаточно прохладным отношением к философии и особенно к философии сознания в ее аналитической версии. Звучали крайне неприятные для философии оценки. «В XX веке ученые длительное время обходили вопрос о сознании

¹ Varela, Francisco J.; Thompson, Evan; Rosch, Eleanor. The embodied mind: Cognitive science and human experience MIT Press, Cambridge, MA, USA. 1991.

² См. *Искусственный интеллект* // Философия: Энциклопедический словарь. Под редакцией А. А. Ивина. М., 2004.

³ См. Солсо Р. Когнитивная психология. М., 1995; Меркулов И. П. Когнитивная наука // Энциклопедия философии и эпистемологии науки. М., 2009.

человека. Некоторые считали, что он слишком отягощен философскими аспектами, трудно поддается экспериментальной проверке или имеет слишком субъективный характер, чтобы подойти к нему с научных позиций»¹. Но в конце концов представители нейронауки пришли к выводу, что изучение сознания требует глубокого междисциплинарного подхода и столь же глубокой философской рефлексии. Особенно когда после огромного количества исследований и блестящих результатов было осознано, что все эти достижения носят эмпирический характер и преимущественно имеют описательный характер. Для того, чтобы получить полноценное объяснение, требуется создание фундаментальной теории мозга, «центрального процессора», «собственника нейронных сетей», как угодно можно его называть. Ввиду такой существенной неполноты знания, авторы указанного издания говорят, что ограничения на использование философии в нейронауках «... в действительности не являются кардинальным препятствием, которого первоначально опасались»².

Отказ от крайностей редукционизма, связанных со сведением психики к нейробиологическим механизмам на уровне описания явлений психики, стремление построить фундаментальные объяснительные теории и отказ от крайностей метафизических рассуждений в философии сознания является характерной чертой продолжающегося синтеза когнитивных нейронаук, философии и искусственного интеллекта. Конкретно-научная и философская методология искусственного интеллекта способна быть основанием дальнейшего развития когнитивной нейронауки, а достижения нейронауки будут, со своей стороны, стимулировать развитие вычислительной техники и систем искусственного интеллекта.

Итак, рассмотрим более подробно, как образуется новая парадигма исследования человеческого мышления, познания и восприятия. Именно такой акцент следует сделать. Речь идет о новой парадигме, которая начинает бурно развиваться с конца 40-х годов 20 века. Если воспользоваться концепцией философа и историка науки Томаса Куна (а его парадигмальная версия описания истории науки будет удобна для презентации теории искусственного интеллекта), то выделение особого этапа в развитии науки, который он называет нормальной наукой, связано со следующими его особенностями.

¹ Баарс Б., Гейдж Н. Мозг. Познание. Разум. Введение в когнитивные нейронауки. Т. 1, М., 2014, с. 54-55.

² Там же, с. 59.

Предполагается наличие фундаментальной теории, которая разрабатывается, детализируется, обрастает массой подтверждающих ее экспериментов. Существует также в этот период множество насущных задач-головоломок, решение которых еще более повышает авторитет фундаментальной теории. Возникает научное сообщество, которое воспроизводит само себя через систему образования. Для воспроизводства научных кадров пишутся фундаментальные учебники и учебные пособия, в которых четко прописываются основные научные результаты, экспериментальные методы и образцы решения задач-головоломок. Если какая-либо головоломка не поддается решению, то виноватыми, как правило, объявляются сами ученые, которые не сумели догадаться, как ее решить. В этом случае такая трудная задача откладывается, считается, что она не влияет на функционирование парадигмы и на роль в ней фундаментальной теории. Только когда таких трудных задач становится достаточно много, они объявляются аномалиями, работа с которыми может привести к научной революции и к образованию новой теории и новой парадигмы. Есть еще одна особенность такого подхода к описанию развития науки. Для обоснования парадигмы и ее фундаментальности привлекаются сведения из истории науки и из истории философской мысли. Кроме того, примеры парадигм и смены фундаментальных теорий в концепции Т. Куна относятся к длительным периодам в истории науки. Куновский анализ не учитывает одновременного существования нескольких параллельных парадигм, что как раз характерно для теории искусственного интеллекта в контексте когнитивной нейронауки. Но парадигмальный анализ очень удобен для рассмотрения не только для исторически растянутых во времени, но и параллельно существующих программ. Характерно, что такое расширенное толкование концепции Т. Куна было очень удачно применено Б. М. Величковским при описании истории научной психологии. Величковский называет современную ситуацию в когнитивной нейронауке методологическим плюрализмом, истоки которого находятся в монадологии Лейбница и философии Христиана Вольфа. «Эта современная монадология, – пишет Б. М. Величковский, – делает возможной совершенно неучтенную Томасом Куном в его теории нормальной науки ... ситуацию, когда разные теории и парадигмы по существу одновременно подходят к разным монадам одной совокупности. Таким образом, прогресс в психологии и когнитивных науках связан как с точностью, так и с плюрализмом, широким видением

ситуации, допускающим существование множество качественно различных «сущностей», не образующих абсолютного единства»¹.

В философии и психологии интерес к исследованию человеческого познания и мышления был настолько устойчивым и постоянным, что говорить о фундаментальной парадигме и ее выделении до середины 19-го века вряд ли уместно. Вялотекущее развитие психологии в рамках философии было прервано с проникновением в этот раздел знания экспериментальной практики. Эмпирическая практика в психологии все более и более становилась критерием оценки гипотез и теорий, сближая психологию с естественными науками и отдаляя ее от философских концепций. Психологические труды и диссертации продолжали включать ссылки на философию во введениях, лишь следуя привычной традиции обоснования новизны авторских подходов. Выход из парадигмы философской психологии и последующую смену парадигм в научной психологии Б. М. Величковский описывает следующим образом (нам это важно, потому что на этапе формирования когнитивной психологии и далее возникают две своеобразные современные парадигмы в теории искусственного интеллекта): «В упрощенном виде картина состоит в следующем. Через каждые 40 лет в психологии меняется ведущая парадигма, причем уже за 20 лет до этого старая парадигма оказывается в состоянии кризиса, а новая начинает прорисовываться в работах будущих классиков. Так, само возникновение научной психологии в период 1860 – 1880 годов происходило на фоне кризисного состояния философской психологии. За этим последовали примерно 20 лет нормальной, в куновском смысле, психологии сознания, доминировавшей на огромных территориях, включая российские центры – Казань, Петербург, Одессу и Москву. Период 1900 – 1920 годов стал временем первого кризиса. Возникновение бихевиоризма и гештальттеории, представители которой были вынуждены постоянно вести диалог с этим направлением, привело к почти полному доминированию психологии поведения. В 1940 – 1960 разразился второй кризис, и на месте бихевиоризма быстро и практически повсеместно утвердилась когнитивная психология. Два десятилетия когнитивизма, с характерными для него компьютерной метафорой, символьным подходом и экспериментами на измерение времени реакции, в свою очередь закончились кризисом 1980 – 2000 годов,

¹ Величковский Б. М. Когнитивная нейронаука. Основы психологии познания. В 2-х т., 2-е издание. М., Юрайт, 2017, т.2, с. 335.

из которого победительницей вышла когнитивная нейронаука, или группа нейронаук – от нейроинформатики, нейропсихологии и нейролингвистики до пока еще несколько более экзотических нейрофилософии и нейроэкономики»¹. Именно в 40-е годы XX века на этапе формирования когнитивной науки возникают практически одновременно и начинают бурно развиваться две парадигмы в теории искусственного интеллекта: парадигма, основанная на символьном вычислительном подходе, и коннекционистская парадигма, основанная на нейросетевом моделировании.

Первая парадигма была более авторитетной, часто именовалась классическим подходом в теории искусственного интеллекта. Что же произошло в этой области науки в середине 20-го века, что заставило по-новому посмотреть на традиционные философские и психологические проблемы исследования мышления и познавательных способностей человека? Принципиально новым для этого периода времени было создание первых электронных вычислительных устройств, которые можно было бы использовать двояким образом. Во-первых, создавать программы для моделирования мышления и познания, и, во-вторых, проверять эти программы на компьютерах. Информатика, рассматриваемая как раздел знания для исследования информации с помощью компьютерных технологий, и программирование легли в основу начинающего приобретать все больший вес символьного вычислительного подхода. Следует обратить внимание на то, что большая часть предварительной работы уже была сделана значительно раньше. Была создана математическая логика и семантика, в основание которых был положен метод формализованных языков. Такой подход предоставил возможность единого описания логики и стал основанием для возникновения гипотезы о возможности формализации человеческого мышления и познания на основе логико-математического символьного метода. К особо важным способностям мышления были отнесены: процесс образования понятий и работа с ними; представление о сущности и роли знания; образование баз данных в различных сферах жизнедеятельности человека; логические способы рассуждения как в классических формах дедукции, индукции и аналогии, так и выходящие за пределы классических описаний, но характерные для естественной способности к рассуждениям; способность выдвигать и проверять гипотезы, анализировать множество гипотез относительно его замкнутости и независимости

¹ Там же, с. 372.

каждой гипотезы от других, способность анализировать условия задачи и решать ее допустимыми методами. Математическая логика и компьютерные науки стали фундаментальными теориями новой парадигмы исследования человеческого интеллекта и познания. Именно в рамках этой парадигмы возникает теория искусственного интеллекта. Можно выделить ряд задач, подобных куновским головоломкам, решение которых, если оно возможно, расширяет область применения теории. Считалось, что существуют две основные фундаментальные проблемы, определяющие спектр приложения теории ИИ. Это – проблема представления знаний и проблема поиска решений задач в пространстве возможных альтернатив. Эти проблемы связаны друг с другом. Нельзя решать задачу без представления ее условий и различных альтернатив ее решения. Многие авторы считают, что обзор возможного множества альтернатив и поиска в нем кандидатов на решение являются методологией и часто применяемой техникой естественного человеческого способа решения поисковых задач. К таким задачам относят, например, игровые эвристики, построение автоматических программ доказательства теорем, создание экспертных систем, понимание естественного языка и связанных с ним проблем обучения, коммуникации и перевода, разработку моделей человеческого разума и *мн. др.*¹. Такой подход к теории и практике искусственного интеллекта был основан на убеждении, что естественное мышление, познание и рассуждение имеют свои законы и принципы действия, которые поддаются исследованию так же, как поддаются изучению законы природы. Открытие и описание этих законов в свою очередь может стать основой математических теорий в области логики, информатики, компьютерных наук и теории искусственного интеллекта. Но в то же время следует отметить, что искусственный интеллект имеет в качестве своей главной предметной области исследование природы человеческого мышления. Однако на ранних стадиях разработки теории искусственного интеллекта он опирался в основном на символические языки математической логики и на компьютерную (вычислительную) метафору, которая очень часто в конкретных исследованиях теряла свое метафорическое значение и понималась уже просто напрямую как способность мозга/разума работать в качестве символьной вычислительной системы. В 1950-80-е годы проводится огромное количество исследований

¹ См. Люгер Дж. Ф. Искусственный интеллект Стратегии и методы решения сложных проблем. М., Издательский дом «Вильямс», 2003, часть 1, гл. 1, § 1.2.

в области моделирования и симуляции человеческого мышления, машинного программирования, поиска решений в пространстве состояний. Формируется представление об использовании компьютерных наук как эмпирической экспериментальной площадки для верификации определенных задач и концепций в области искусственного интеллекта. В конце концов оформляется концепция, которую можно назвать классическим символьным подходом в теории искусственного интеллекта. Его суть выражает гипотеза о физической символьной системе, обобщенное описание которой было представлено Алленом Ньюэллом и Гербертом Саймоном. Полная версия этой гипотезы такова:

«Физическая система проявляет разумное в широком смысле поведение тогда и только тогда, когда она является физической символьной системой.

Достаточность означает, что разумность может быть достигнута каждой правильно организованной физической символьной системой.

Необходимость означает, что каждый агент, проявляющий разумность в общепринятом смысле, должен являться физической символьной системой. Необходимое условие этой гипотезы требует, чтобы любой разумный агент, будь то человек, инопланетянин или компьютер, достигал разумного поведения путем физической реализации операций над символьными структурами.

Разумное в широком смысле поведение (*general intelligent action*) означает действия, характерные для поведения человека. Фактически ограниченная система ведет себя соответственно своим целям, приспосабливаясь к требованиям окружающей среды¹».

Использование символьного подхода привело к особому способу описания мира и представлению знания о нем. Этот подход был наследником мощной рационалистической традиции. Еще Г. В. Лейбниц мечтал об универсальном методе проверки рассуждений, основанном на формальном символическом представлении знаний. «Точность, – писал он в «Теодицее», – нас затрудняет, и правила представляются нам лишь ребячеством. Вот почему общепринятую логику (однако же почти достаточную для исследования рассуждений, претендующих на достоверность) считают принадлежностью одних лишь школьников и нисколько не заботятся о логике, которая должна взвешивать степень вероятного и которая столь необходима

¹ Там же, с. 782.

при важных решениях. Таким образом, истинно то, что наши ошибки большей частью проистекают от отвращения к мышлению или от недостатка искусства мыслить...»¹

Стоит особо отметить также книгу Джорджа Буля «Исследование законов мышления, на которых основываются математические теории логики и вероятностей» (1854), оказавшую значительное влияние на создание новой математической логики и будущей вычислительной парадигмы в когнитивных науках и теории искусственного интеллекта. Буль считал, что описанная им алгебра логики представляет собой законы мыслительной деятельности. И именно эта алгебра должна служить фундаментальным базисом математической логики. Именно так! Математическая логика – искусственная теоретическая концепция – основана на открытых Булем, как он полагал, законах естественного человеческого мышления. Еще раз подчеркиваем, что Джордж Буль считал свою алгебру логики не прикладной теорией для описания логических способностей человека, а теорией, основанной на опыте естественных способов рассуждения людей. В свою очередь математическая логика тоже опирается на естественные законы мышления.

Именно такая идеология лежит в основе гипотезы о физической символической системе А. Ньюэлла и Г. Саймона: любая физическая система является интеллектуальной, если она является символической, организованной по правилам вычислительных операций. Поведение такой системы аналогично рациональному поведению человека, включает цель и намерения. Такая система способна адаптироваться к условиям окружающей среды.

Что касается коннекционистской парадигмы, то она изначально возникает в нейронауке и в теории искусственного интеллекта, т.е. в конкретно-научных и прикладных технических исследованиях. Джеймс Гарсон в статье «Коннекционизм», опубликованной в Стэнфордской энциклопедии по философии, дает следующее определение коннекционизма: «Коннекционизм – это движение в когнитивной науке, которое надеется объяснить интеллектуальные способности человека при помощи использования искусственных нейронных сетей ... Нейронные сети являются упрощенными моделями мозга... Эксперименты на моделях такого рода продемонстрировали способность к обучению таким навыкам, как распознавание лиц, чтение и обнаружение простой грамматической структуры. Философы

¹ Г. В. Лейбниц. Собр. соч. Т. 4, с. 95

заинтересовались коннекционизмом, потому что он обещал обеспечить альтернативу классической теории разума: широко распространенному мнению о том, что разум является аналогом цифровой вычислительной процедуры в символическом языке. В какой степени коннекционистская парадигма представляет собой альтернативу классике, стало предметом горячих дискуссий в последние годы (последняя версия цитируемой работы относится к 2007 г. – В. К.)». Заметим, что философы обратили внимание на коннекционизм позже того, как он уже возник в конкретных науках и технических приложениях. В. А. Лекторский о времени возникновения коннекционизма пишет следующее: «Развитие современных когнитивных наук ... исходит из того, что познание может и должно быть понято как совокупность процессов переработки мозгом или каким-то другим устройством – естественным или искусственным – информации, поступающей из внешнего мира... В процессе развития когнитивных наук за последние 30 лет был собран огромный эмпирический материал. Seriously менялись и общие представления относительно когнитивных процессов; в частности, на смену идеям о существовании «языка мозга», ментальных репрезентаций и алгоритмических способов переработки когнитивной информации пришли идеи т.н. коннекционизма, а затем концепция динамических когнитивных систем»¹. Это важно, потому что коннекционизм появился в физиологии мозга в XIX веке как альтернатива ошибочным теориям, которые приписывали отдельным участкам мозга конкретные функции и которые пытались объяснить интеллектуальные различия людей и отклонения от нормы размерами головного мозга и его зон. Эти теории не объясняли, как и почему отдельные участки головного мозга выполняют свойственные им функции. Коннекционизм попытался восполнить этот пробел. Профессор компьютерной нейробиологии Массачусетского технологического института Себастьян Сеунг писал о коннекционизме следующее: «Эта теория также берет начало в XIX веке. Подход этот в идейном смысле более амбициозен: в его рамках делается попытка объяснить, как же, собственно, работают участки

¹ Лекторский В. А. Познание и сознание: дискуссия анти-реализма и реализма в современной эпистемологии // Проблема сознания в философии и науке. М., «Канон+» РООИ «Реабилитация», 2009, с. 64–65. К интересующему нас вопросу относятся еще две публикации этого же автора: Лекторский В. А. Философия, когнитивная наука и искусственный интеллект // Искусственный интеллект: междисциплинарный подход / Под ред. Д. И. Дубровского, В. А. Лекторского. М., 2006; Лекторский В. А. Философия и исследование когнитивных процессов // Когнитивный подход: философия, когнитивная наука, когнитивные дисциплины / Под ред. В. А. Лекторского. М., 2007

мозга. Коннекционисты рассматривают зону мозга не как некую элементарную ячейку, а как сложную сеть, которая состоит из множества нейронов. Связи в этой сети организованы таким образом, чтобы нейроны могли предпринимать коллективные действия того или иного рода, которые и лежат в основе нашего восприятия и наших мыслей. Организацию этих связей можно изменять, приобретая новый опыт, что и позволяет нам учиться и запоминать новое. Организацию этих связей формируют и гены ... так что генетическое воздействие на ум и сознание также можно объяснить в рамках предлагаемой теории. Эти идеи могут показаться весьма многообещающими, но следует иметь в виду: их никогда не подвергали корректной экспериментальной проверке, которая давала бы убедительные результаты. Коннекционизм, несмотря на свою привлекательность с интеллектуальной точки зрения, так и не сумел стать настоящей наукой, поскольку нейробиологам всегда не хватало технологий, которые позволили бы должным образом картировать межнейронные связи». Еще раз обращаю внимание на то, что в данной статье речь идет о коннекционистском подходе в теории и технических разработках искусственного интеллекта¹.

Д. А. Поспелов в своей знаменитой статье «Десять «горячих точек» в исследованиях по искусственному интеллекту» (опубликованной в сборнике «Интеллектуальные системы». М., 1996. Т. 1, вып. 1-4. См. [17]) девятой «горячей точкой» называет «сетевые модели» и сообщает буквально следующее: «Интеллектуальные системы, основанные на правилах (продукциях), принесли не только радость решения ряда важных задач, но и породили сомнения в том, что именно они призваны остаться основными моделями представления знаний в интеллектуальных системах. Многочисленные дискуссии 80-х годов, проводившиеся специалистами в области ИИ по этому поводу, привели к укреплению сетевой парадигмы, несколько отодвинутой в сторону триумфальным выходом на сцену продукционных моделей. И хотя исследования в области семантических сетей, каузальных сетей и сетей другого типа продолжались, они были малочисленными и не слишком продуктивными».

Но к концу 80-х годов сетевые модели стали развиваться более быстрыми темпами. Этот процесс совпал с пробуждением интереса к давно забытым нейронным архитектурам (выделено мною – В. К.),

¹ Сеунг С. Коннектом. Как мозг делает нас тем, что мы есть. 2-е изд. М., Лаборатория знаний, 2016, с. 20. На смену коннекционистскому подходу, по мнению С. Сеунга, приходит коннектомика. Об этом его книга.

появлением транспьютерных систем и нейрокомпьютеров, а также с возвращением к работам, опирающимся на эволюционные модели и эволюционное программирование. Возник определенный бум, который был даже окрещен неodarвинизмом.

Если к концу первого этапа развития сетевых моделей (в основном в виде нейронных многослойных систем типа персептронов) наступило разочарование в их возможностях и простоте их аппаратной реализации, то в 80-х годах эти сомнения были отброшены. Комплекс исследований в этой области так возрос, что произошло практическое отпочкование специалистов, работающих в области сетевых моделей, от основного ядра тех, кто причисляет себя к искусственному интеллекту. У «сетевиков» появились свои журналы, они стали проводить свои симпозиумы и конференции и формировать свою терминологию. Этот разрыв нарастает, что по-видимому, приведет к возникновению двух наук, связанных с построением интеллектуальных систем. Одна из них будет по-прежнему опираться на уровень ментальных (информационных) представлений, а другая – на уровень структурной организации (по типу нервных тканей), порождающей нужные решения. Во всяком случае в 90-е годы вряд ли можно ожидать спад интереса к сетевым моделям и многочисленным нерешенным проблемам, связанным с их построением и функционированием». Следует заметить, что эта небольшая работа нашего выдающегося отечественного ученого в области логики, информатики, компьютерных наук и искусственного интеллекта очень четко представляет историю искусственного интеллекта и взаимоотношение двух ее основных парадигм. Коннекционизм как парадигмальное направление в искусственном интеллекте возникает в 80-е годы XX века. Все куновские признаки парадигмы четко определились. Об этом же времени окончательного образования коннекционизма свидетельствует Дэниел Деннет. Интересно отметить, что эти результаты были получены абсолютно независимо друг от друга в нашей стране и США. Коннекционизм этого периода можно назвать «нормальной наукой». Д. Деннет со ссылкой на статью Фельдмана и Балларда «Коннекционистские модели и их свойства», вышедшую в 1982 году, утверждает, что к этому времени уже возникли исследовательские стратегии, которые можно было бы обобщенно называть коннекционизмом. Предыстория этого направления, разумеется, исчисляется солидным интервалом времени. Однако интересно, что буквально через семь лет (т.е. к тому времени, о котором писал

Д. А. Поспелов) научное сообщество коннекционистов значительно расширяется за счет появления новых исследовательских программ PDP (parallel distributed processing) и нейросетевого моделирования¹. Деннет показывает, что данное сообщество не является единым и однородным, представители разных исследовательских программ могут опираться на разные принципы и исповедовать разные идеологии. Но эти дискуссии ведутся внутри одного лагеря ученых. Они представляют разные исследовательские программы внутри одной коннекционистской парадигмы.

Примерно в это же время появляется знаменитая статья Дж. Фодора и З. Пылишина «Коннекционизм и когнитивная архитектура: критический анализ»². Многие исследователи считают, что критика коннекционизма этими авторами, основанная на попытке доказать, что отношение между разумом и структурой мозга на когнитивном уровне не является нейросетевым, привела к дебатам, которые были подхвачены философами. Появилась особая стадия коннекционизма, которую стали называть «новым коннекционизмом». На этой стадии коннекционистские конкретно-научные исследования стали особым предметом обсуждения в философии когнитивной науки и философии сознания³. «Новый коннекционизм» – это направление в философии, когда предметом исследования стали сознание и источники познания, а основной методологией – искусственные нейросетевые модели. Некоторые авторы пытались обвинить философов в том, что они, внедрившись в конкретно-научную проблематику, внесли путаницу и даже в определенном смысле мифологизировали коннекционистские исследования. Так, в частности, Иштван Беркли в работе «Некоторые мифы коннекционизма» формулирует четыре таких «мифа».

Первый миф состоит в том, что коннекционистские системы являются в некотором смысле нейронными или похожими на мозг. В другой формулировке: коннекционистские системы обладают свойствами, похожими на свойства мозга. По мнению И. Беркли, эта идея

¹ *Dennett, Daniel C.* Mother nature versus the walking encyclopedia. In William Ramsey, Stephen P. Stich & D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. 1991.

² *Fodor, J. and Pylyshyn, Z.* «Connectionism and Cognitive Architecture: A Critical Analysis», in *Cognition* 1988, 28, pp. 3-71. Рус. Пер.: Фодор Дж., Пылишин З. Коннекционизм и когнитивная структура: критический обзор. В кн. *Язык и интеллект*. М.: Прогресс, 1995.

³ *Berkeley Istvan S. N.* Some Myths of Connectionism. 1997. *Philosophy*, The University of Southwestern Louisiana.

была вброшена в философскую литературу в редакции Э. Кларка («коннекционистская архитектура подобна структуре мозга»)¹. Бехтель и Абрахамсен, кроме того, не просто принимают аналоговую сетевую модель, но и связывают с ней подъем нового коннекционизма². Но наиболее четко этот миф был выражен Полом Черчлендом. Беркли приводит следующую цитату из работы Пола Черчленда: «Сети с искусственными единицами пытаются исследовать природные нейроны»³. Даже Д. Деннет ссылается на коннекционистские структуры с нейроподобными элементами. Интересно отметить замечание Беркли, что некоторые авторы, которые даже не являются сторонниками коннекционизма, принимают эту аналогию. Беркли отмечает, что имеется и другая точка зрения, не нашедшая отражения в философии: «Тем не менее, в технической литературе часто открыто признается, что существуют радикальные несходства между биологическими и коннекционистскими системами (смотри, например, работу Крика и Асанума 1986 года⁴). К сожалению, эти благоприятные замечания редко прокладывают свой путь в философскую литературу»⁵.

Второй миф: коннекционистские системы работают в реальном времени. Происходит явный перенос особенностей работы мозга человека на нейронную модель. Несмотря на то, что человеческий мозг работает, «вычисляет» медленнее, чем компьютер, он способен производить обработку данных в реальном времени за счет параллелизма алгоритмической работы элементов в нейронной сети. Преимущества быстрой последовательной компьютерной обработки нивелируются за счет PDP. Эта концепция разделяется многими исследователями и очень четко определяется Д. Румельхартом: «Хотя мозг имеет медленные компоненты, их очень много ... Вместо организации вычислений при помощи множества последовательных

¹ Clark, A. *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*, MIT Press (Cambridge, Mass.), 1989, p. 4.

² Bechtel, W. and Abrahamsen, A. *Connectionism and the Mind*, Basil Blackwell (Cambridge, Mass.), 1991, p. 17.

³ Churchland, P. M. *The Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press (Cambridge, Mass.). 1989, p. 160.

⁴ Crick, F. and Asanuma, C. (1986), «Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex», in Rumelhart, McClelland et al. (1986: Vol. 2., pp. 333-371). Интересно, что эта статья была опубликована в сборнике, который называют «Библией» коннекционизма.

⁵ Berkeley Istvan S. N. *Some Myths of Connectionism*. 1997. Philosophy, The University of Southwestern Louisiana.

шагов, как мы делаем в системах с быстрыми шагами, мозг совершает множество элементов обработки совместно и параллельно осуществляет свою деятельность»¹.

Выдвижение положения, что параллельно распределенная обработка «сжимает» интервал времени до примерно такого же интервала, который требуется для совершения одинаковой работы последовательно устроенными компьютерными программами, объявляется Беркли мифом. Он пишет, что аргумент «сжатия» времени за счет параллельной обработки «... не следует рассматривать как оказание поддержки утверждению, что коннекционистские системы согласуются с реальными ограничениями времени при обработке. Это, по крайней мере, в качестве общего требования для коннекционистских систем есть просто еще один коннекционистский миф»².

Если два предыдущих мифа основаны на выдвижении положения о сходстве коннекционистских систем с биологическими организмами, то третий миф связан с попыткой доказать предпочтительность коннекционистского подхода по отношению к неконнекционистским системам. Коннекционизм предоставляет возможность работы с неполными или несовершенными входными данными. Пол Черчленд отмечает: «Вы можете распознать фотографию лица вашего лучшего друга в любой из широкого спектра поз менее чем за половину секунды. Но такая ориентировка до сих пор ускользает от лучших программ распознавания образов, которые имеются в наличии. ... Даже сильно упрощенные проблемы распознавания образов на самом деле являются очень трудными для неконнекционистских систем»³. На самом деле никакие системы в условиях неполноты или смутности информации не имеют преимущества друг перед другом.

Четвертый миф связан с обучаемостью коннекционистских систем. Они умеют переносить навыки, полученные в ходе обучения (это интерпретируется как практический опыт), на новые случаи (интерпретируется как способность к обобщению) лучше, чем неконнекционистские системы.

¹ *Rumelhart, D.* «The Architecture of Mind: A Connectionist Approach», in Posner, M., *Foundations of Cognitive Science*, MIT Press (Cambridge, Mass.) 1989, p. 135.

² *Berkeley Istvan S. N.* Some Myths of Connectionism. 1997. Philosophy, The University of Southwestern Louisiana.

³ *Churchland, P. M.* (1988), *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, MIT Press (Cambridge, Mass.), p. 120. Цитируется по указанной работе Беркли.

Приведенный мною пример критики коннекционизма является типичным явлением для дискуссии между классиками и коннекционистами в 1980-90-х годах. Но к настоящему времени актуальность такого рода дискуссий утрачена. Проблема обучения стала одной из главных и трудных проблем для обеих парадигм искусственного интеллекта. Дж. Люгер отмечает: «Обучение является «крепким орешком» искусственного интеллекта. Важность обучения, тем не менее, несомненна, поскольку эта способность является одной из главных составляющих разумного поведения»¹. В цитируемой книге Дж. Люгера описаны многочисленные методы машинного обучения на основе символического, нейросетевого и эмерджентного подходов до начала XXI века. Проблема обучения стала общей проблемой для всех парадигм искусственного интеллекта. Разумеется, проблема обучения не является единственной общей и трудной проблемой для обеих парадигм. Такими проблемами являются, например, проблемы распознавания образов, эвристического поиска, решения задач и принятия решений, создания экспертных систем, источников и баз знаний, представления знаний, понимания естественного языка и мн. др.

Многообразие решений, концепций и методов подводит к принятию обобщенного определения искусственного интеллекта как области теоретического и прикладного исследования принципов решения интеллектуальных задач. Опираются ли эти решения на аналогию с естественным разумом или его имитацию, имеет принципиальное значение, хотя бы потому, что существуют программы, выполняющие актуальные интеллектуальные задачи и не опирающиеся на принципы человеческого разума², т.е. не являющиеся биологически сходными с естественным разумом. Искусственный интеллект, как это было показано В. К. Финном, – это «направление исследований компьютерной науки, в котором разрабатываются средства имитации и усиления умственной активности человека»³. И это действительно, верно. Многие исследователи говорят о непродуктивности продолжения дискуссии между классиками и коннекционистами. Так, например, Б. М. Величковский, подводя итог дискуссии о природе обра-

¹ Люгер Дж. Ф., цит. Соч., с. 50.

² См. Иваницкий А. М. Проблема сознания и физиология мозга // Проблема сознания в философии и науке. М., «Канон+» РООИ «Реабилитация», 2009. О проблеме искусственного интеллекта с. 390-394.

³ Финн В. К. Искусственный интеллект: Методология, применения, философия. М., URSS. 2011.

за, пишет: «Спор о том, являются ли образы аналого-перцептивными или логико-лингвистическими репрезентациями, по-видимому, теряет свою недавнюю актуальность, становясь достоянием истории»¹.

В заключение хотелось бы сказать, что поскольку искусственный интеллект основан на изучении человеческого разума, а научных интерпретаций разума очень много, то и каждая из теорий искусственного интеллекта может претендовать в свою очередь только на определенную интерпретацию. Вопрос о ее научном статусе решает практическая способность адекватно выполнить поставленную интеллектуальную задачу средствами, имитирующими или даже превосходящими естественные способности разума. Искусственный интеллект занимается решением интеллектуальных задач. Решить глобальную задачу создания искусственного коррелята сознания, каким бы он ни был, символьным или нейросетевым, пока еще никому не удалось. Причиной, объясняющей такое положение дел, является отсутствие фундаментальной теории, связывающей биологические свойства мозга со свойствами субъективной реальности. Роберт Бёртон в своей книге, которая в русском переводе получила название «Разум vs мозг. Разговор на разных языках», писал: «Верить в то, что мы можем найти нейрональный коррелят сознания, все равно, что верить, что рассмотрения нейронов и их связей достаточно для характеристики поведения, организованного на более высоких уровнях сложности. С моей точки зрения, это представляет собой ошибку категоризации, которую нельзя преодолеть путем улучшения техники и методов. Только когда мы перекинем мост через провал между фундаментальной физиологией и высокоуровневыми эмерджентными свойствами и выразим это понимание в физиологических терминах, у нас появятся теоретические шансы описать нейрональные корреляты сознания». Скептицизм Бёртона не является абсолютным. Он оставляет надежду на создание фундаментальной теории, которая перекинет «мост через провал». Таких теоретических набросков существует уже несколько. Но это уже другая история.

¹ Величковский Б. М. Когнитивная нейронаука. Основы психологии познания. В 2-х т., 2-е издание. М., Юрайт, 2017, т. 2, с. 294.

ЛИТЕРАТУРА

1. Баарс Б., Гейдж Н. Мозг. Познание. Разум. Введение в когнитивные нейронауки. В 2-х т. 1, М., 2014
2. Бёртон Р. Разум vs мозг. Разговор на разных языках. М., Изд-во «Э», 2016.
3. Величковский Б. М. Когнитивная нейронаука. Основы психологии познания. В 2-х т., 2-е издание. М., Юрайт, 2017.
4. Дрейфус Х. Л., Дрейфус С. И. Создание сознания vs. Моделирование мозга: Искусственный интеллект вернулся на точку ветвления // Аналитическая философия: становление и развитие. Антология. Общая редакция и составление А. Ф. Грязнова. М., 1998.
5. Дубровский Д. И. Проблема «Сознание и мозг». Теоретическое решение. М., «Канон+» РООИ «Реабилитация», 2015.
6. Иваницкий А. М. Проблема сознания и физиология мозга // Проблема сознания в философии и науке. М., «Канон+» РООИ «Реабилитация», 2009
7. Искусственный интеллект // Философия: Энциклопедический словарь. Под редакцией А. А. Ивина. М., 2004.
8. Кун Т. Структура научных революций. Прогресс, М., 1977.
9. Кузнецов В. Г. Парадигмальная философия науки Томаса Куна // Философия. Под ред. А. Ф. Зотова, В. В. Миронова, А. В. Разина. Академический проект, М., 2004.
10. Лейбниц Г. В. Собр. соч. в 4-х т., т. 4, М., 1984.
11. Лекторский В. А. Познание и сознание: дискуссия антиреализма и реализма в современной эпистемологии // Проблема сознания в философии и науке. М., «Канон+» РООИ «Реабилитация», 2009.
12. Лекторский В. А. Философия, когнитивная наука и искусственный интеллект // Искусственный интеллект: междисциплинарный подход / Под ред. Д. И. Дубровского, В. А. Лекторского. М., 2006.
13. Лекторский В. А. Философия и исследование когнитивных процессов // Когнитивный подход: философия, когнитивная наука, когнитивные дисциплины / Под ред. В. А. Лекторского. М., 2007.
14. Люгер Дж. Ф. Искусственный интеллект. Стратегии и методы решения сложных проблем. М., Издательский дом «Вильямс», 2003.
15. Меркулов И. П. Когнитивная наука // Новая философская энциклопедия. Т. 2, М., 2001.
16. Меркулов И. П. Когнитивная наука // Энциклопедия философии и эпистемологии науки. М., 2009.
17. Поспелов Д. А. Десять «горячих точек» в исследованиях по искусственному интеллекту // Интеллектуальные системы. М., 1996. Т.1, вып. 1-4.
18. Сеунг С. Коннектом. Как мозг делает нас тем, что мы есть. 2-е изд. М., Лаборатория знаний, 2016.
19. Солсо Р. Когнитивная психология. М., 1995.

20. Финн В. К. Искусственный интеллект: методология, применения, философия. М., URSS. 2011.
21. Фодор Дж., Пилишин З. Коннекционизм и когнитивная структура: критический обзор. // Язык и интеллект. М.: Прогресс, 1995..
22. Фрим Крис. Мозг и душа. Как нервная деятельность формирует наш внутренний мир. М., 2015.
23. Bechtel, W. and Abrahamsen, A. Connectionism and the Mind, Basil Blackwell (Cambridge, Mass.), 1991.
24. Berkeley Istvan S. N. Some Myths of Connectionism. 1997. Philosophy, The University of Southwestern Louisiana.
25. Churchland, P. M. Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind, MIT Press (Cambridge, Mass.). 1988.
26. Churchland, P. M. The Neurocomputational Perspective: The Nature of Mind and the Structure of Science, MIT Press (Cambridge, Mass.). 1989.
27. Clark, A. Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing, MIT Press (Cambridge, Mass.), 1989.
28. Crick, F. and Asanuma, C. «Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex», in Rumelhart, McClelland et al. (1986: Vol. 2., pp. 333-371).
29. Dennett, Daniel C. Mother nature versus the walking encyclopedia. In William Ramsey, Stephen P. Stich & D. Rumelhart (eds.), Philosophy and Connectionist Theory. 1991.
30. Rumelhart, D., McClelland, J. and The PDP Research Group. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, (2 Vols.), MIT Press (Cambridge, Mass.). 1986.
31. Rumelhart, D. «The Architecture of Mind: A Connectionist Approach», in Posner, M., Foundations of Cognitive Science, MIT Press (Cambridge, Mass.) 1989.
32. Fodor, J. and Pylyshyn, Z. «Connectionism and Cognitive Architecture: A Critical Analysis», in Cognition 1988, 28, pp. 3-71.
33. Varela, Francisco J.; Thompson, Evan; Rosch, Eleanor. The embodied mind: Cognitive science and human experience MIT Press, Cambridge, MA, USA. 1991.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НАУКИ О МОЗГЕ – МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ

ARTIFICIAL INTELLIGENCE AND SCIENCES OF BRAIN – METHODOLOGICAL PROBLEMS

库姿妥册扶 «人工智能寻找新范式的理论»

**Кузнецов
Олег Петрович**

*Председатель совета Российской ассоциации
искусственного интеллекта, доктор технических
наук, профессор, заведующий лабораторией
Института проблем управления РАН, г. Москва*

Kuznetsov Oleg P.

*PhD, Dr. Sci. (Engineering), head of a department in
Trapeznikov Institute of Control Sciences of RAS, Moscow*



Аннотация. Рассматривается история искусственного интеллекта. Отмечается недостаточность компьютерной парадигмы и подчеркивается необходимость более тесной связи искусственного интеллекта с науками о мозге. Обсуждаются методологические принципы построения теории мозга.

Ключевые слова: история искусственного интеллекта, методология, науки о мозге.

Annotation. The history of artificial intelligence is considered. An insufficiency of computer paradigm is acknowledged. A necessity of close relationship with the brain sciences is emphasized. The methodological principles for development of brain theory is discussed.

Keywords: history of artificial intelligence, methodology, science of the brain.

О т других когнитивных наук искусственный интеллект отличается целевой установкой. В отличие от других когнитивных наук – наук о мозге – искусственный интеллект относится к компьютерным наукам. Он исследует не мозг как таковой, а средства решения интеллектуальных задач, то есть задач, которые человек решает не с помощью алгоритмов, а путем использования опыта, знаний, сообразительности и рассуждений.

Более пятидесяти лет в искусственном интеллекте безраздельно господствовала компьютерная парадигма. Ее основные положения:

мозг – это биокомпьютер, и его когнитивные механизмы надо рассматривать как механизмы переработки информации;

электронные компьютеры эффективнее биологических компьютеров: они работают в миллион раз быстрее, не ошибаются, не устают и т.д. ;

следовательно, мозг изучать не надо; искусственный разум будет создан другим путем – путем создания компьютерных программ.

Такой подход называют феноменологическим. Он идет от теста Тьюринга, который сам Тьюринг в знаменитой работе «Может ли машина мыслить?» [1] назвал игрой в имитацию, предложив считать машину мыслящей, если при общении с ней ее нельзя отличить от человека.

Один из лидеров советской кибернетики Виктор Михайлович Глушков в интервью «Литературной газете» 1 января 1976 года на вопрос, можно ли создать искусственный мозг, сказал: «Да, это вообще не вопрос. Безусловно, можно. И это все произойдет еще до начала XXI века». В качестве аргументации была та же идея: не обязательно смотреть, как мозг устроен внутри, важно, чтобы результат был тот же самый.

Интересно отметить, что именно в то же время (в 1956 г.), когда искусственный интеллект был объявлен самостоятельной наукой и прочно встал на позиции феноменологической (компьютерной) парадигмы, возникла когнитивная психология, лидеры которой заявили о категорическом отказе от основных положений бихевиоризма – аналога феноменологического подхода в психологии – и о намерении изучать внутренние механизмы мозга.

Итак, долгое время контакты между искусственным интеллектом и науками о мозге были минимальными. Среди немногих ученых, которые понимали важность таких контактов и старались их поддерживать, следует отметить Дмитрия Александровича Поспелова

(организатора и первого председателя Российской Ассоциации искусственного интеллекта). Тем не менее, успехи искусственного интеллекта были значительными. Помимо решения конкретных интеллектуальных задач, главные достижения искусственного интеллекта заключались в том, что он, во-первых, накопил огромный опыт формализации и моделирования познавательных структур и процессов; и, во-вторых, предложил компьютерные аналоги всех человеческих интеллектуальных средств, то есть опыта, знаний, сообразительности и рассуждений.

Опыт – это машинное обучение.

Знания – это инженерия знаний [2]: различные способы извлечения знаний, их представления в виде продукций, семантических сетей, онтологий, а также хранения и использования знаний в виде баз знаний.

Сообразительность – это разнообразные эвристические методы.

Рассуждения – это логические методы: дедуктивная логика и метод резолюций [3, 4], индуктивная логика и ДСМ-метод [5], а также различные виды неклассических логик [4].

К 90-м годам стало ясно, что, несмотря на успехи искусственного интеллекта и многократно выросшую эффективность вычислительной техники, по-прежнему остаются очень важные задачи, которые мозг решает эффективнее компьютера:

быстрая обработка образов: категоризация, узнавание в разных ракурсах на основе сходства (а не тождества);

целостное, «гештальтное» восприятие; восстановление целого по части; быстрое узнавание несходства («что-то не так»);

определение релевантности, отделение существенного от несущественного;

быстрый доступ к нужному содержанию (ассоциативный поиск);

быстрые рассуждения на основе схем [6, 7].

Появилось понимание того, что мозговые и компьютерные механизмы обработки информации принципиально различны: «то, что просто для компьютера, сложно для мозга, и наоборот» [8]. Отсюда следует, что разгадка мозговых механизмов решения вышеуказанных задач и их электронная реализация должны привести к революционным изменениям в информационных технологиях. Внутри искусственного интеллекта началось медленное, но неуклонное движение в сторону наук о мозге, связанное с усилением интереса к изучению когнитивных механизмов мозга.

С этим пониманием возникло осознание того, что создание искусственного разума – дело гораздо более сложное, чем это казалось в начале пути. Если В. М. Глушков в 1976 г. предсказывал это создание к 2000 г., т.е. через 25 лет, то в 2000-м году считалось, что понадобится еще не менее 25 лет. Эту же цифру называют и сейчас. Создается впечатление, что 25 лет – это что-то похожее на расстояние до линии горизонта: сколько ни идешь, оно остается прежним.

В связи с этим возникла дискуссия по проблеме того, до какого уровня изучения мозга надо спускаться? Достаточно ли уровня поведения, то есть психологии (аналогично тому, что программисту не нужно знать, как устроены электронные схемы компьютера), или все же нужно спускаться на нейрофизиологический уровень?

На мой взгляд, знание нейрофизиологического уровня необходимо не только для изучения собственно мозга, но и для решения внутренних стратегических задач ИИ – создания продвинутых информационных технологий. Для того, чтобы пользоваться устройством, программой или технологией, достаточно знать, грубо говоря, на какие кнопки нажимать. Однако для того, чтобы воспроизвести эти объекты и их функциональные свойства, надо знать, как они устроены.

Более того, возможно, придется спускаться и на биохимический, и на генетический уровень.

Одновременно с этим движением возникло встречное движение когнитивных наук, у которого были свои причины. Главная из них состоит в том, что при огромном количестве накопленных экспериментальных данных в науках о мозге (и, в первую очередь, в нейрофизиологии) остро не хватает теоретических подходов.

Об этом свидетельствует, в частности, Джефф Хокинс в книге [8]: «Мои поиски удовлетворительной теории по работе мозга оказались безрезультатными. Я пришел к выводу, что ни у кого нет идей, как действительно работает мозг. Не было даже плохой или недоказанной теории; не было вообще никакой. Это было странно. Например, никто точно не знал, почему вымерли динозавры, но было изобилие теорий, о которых можно было прочитать. Но относительно работы мозга не было ничего похожего».

Об отсутствии теории мозга и необходимости создания постоянно говорит К. В. Анохин в своих докладах последних лет [9, 10]. В одном из докладов на семинаре НСМИИ в 2014 г. он говорил о головоломках и тайнах в науке. Головоломка – это поставленная задача,

про которую непонятно, как ее решать. А тайна – это ситуация, когда задача не только неясна, но и непонятно, «как про нее думать». Науки о мозге пребывают именно в таком состоянии. Оно означает, что а) нет языка, то есть системы понятий, в терминах которой можно обсуждать проблемы и строить гипотезы; б) соответственно, нет средств построения гипотез, то есть верифицируемых моделей.

Теоретический голод привел нейрофизиологов десять лет назад к «сетевому буму», когда они обнаружили, что современная теория сложных сетей [11 – 13] отлично им подходит. Эта теория впервые дала нейрофизиологам словарь и систему точных понятий для описания нейронных структур как анатомических (структурных), так и функциональных. Стал весьма популярным термин «коннектом». Этим термином принято сейчас называть нервную сеть мозга, хотя имеются разночтения в трактовке этого термина.

Тем не менее, этого языка оказалось недостаточно для построения теории мозга. Он хорошо подходит для описания структурных свойств нервной системы, но не имеет средств для описания функциональных свойств. Язык коннектома – это язык теории графов. Вершины графа – это всего лишь узлы, пассивно передающие возбуждение по сети. Они не наделены функциональностью, т.е. не способны преобразовывать информацию. В нем нет средств описания семантики (содержания деятельности мозга) и динамики (развертывания когнитивных процессов во времени). Поэтому теорию мозга еще предстоит построить.

Как ее строить?

Ситуация в нейрофизиологии, характеризующаяся тем, что при огромном массиве экспериментальных данных теория фактически отсутствует, является выразительным подтверждением попперовской критики индуктивизма. Теорию не получается построить снизу, путем обобщений. Язык теории не может совпадать с языком эксперимента. Более того, в развитых науках именно теория подсказывает направления экспериментальных исследований, которые должны подтверждать или опровергать положения теории. Как сказал когда-то А. Эйнштейн, «Лишь теория решает, что мы ухищряемся наблюдать».

Такой язык, на наш взгляд, следует строить на основе понятий информатики с использованием математического аппарата теории графов, сложных сетей и гиперсетей, а также алгоритмических моделей (логических сетей, конечных автоматов и др.). Именно

искусственный интеллект владеет этими средствами и методологическим опытом построения формальных моделей.

Концепция когнитивного К. В. Анохина [9, 10] представляет собой важный шаг на этом пути. В то же время помимо далеких целей типа объяснения того, что такое сознание, необходимо особое внимание уделить ближним целям. Такой ближней целью, как мне представляется, должно быть создание минимальной модели.

Минимальная модель – это совокупность принципов, которая предлагает объяснение ближних проблем, дает адекватную систему понятий, покрывающую всю область; в ее терминах формулируются утверждения и гипотезы. Задача минимальной модели – не в том, чтобы всё объяснить, а в том, чтобы дать языковые средства для обсуждения проблем. При дальнейшем развитии модель пополняется, но основные ее принципы и понятия остаются верными.

Все сложившиеся науки имеют минимальные модели, хотя их развитие продолжается, и ни в одной из них все проблемы не решены. Это механика Ньютона для физики, планетарная модель атома и таблица Менделеева для химии, эволюционная концепция и генетика для биологии, машина Тьюринга для вычислительных наук.

В науке о мозге на роль минимальной модели претендовали: модель компьютера (компьютерная парадигма), разные модели нейронных сетей, а теперь еще и коннектом. Все они на эту роль не годятся прежде всего потому, что они не содержат семантических понятий.

Идея когнитивного как раз семантическая, и она имеет все шансы стать частью (возможно, основой) минимальной модели. Но для этого она должна стать более конструктивной, не претендовать на решение дальних проблем и, напротив, более четко формулировать ближние проблемы, среди которых я вижу в первую очередь построение точных моделей основных когнитивных механизмов мозга в терминах обработки информации: кодирования, хранения, доступа, считывания и т.д. При этом теория мозга даже в далекой перспективе не должна строить модель конкретного мозга с точностью до нейрона. Однако она должна дать набор параметров, в терминах которых можно говорить о различиях мозга у разных людей. Очень много известно об анатомических параметрах, однако какую роль они играют в когнитивных процессах, во многом остается неизвестным.

Сложившаяся теория опирается на парадигму – установившееся представление о том, что имеется общепринятая система понятий и запретов, в терминах которой можно задать все нужные вопросы и

получить все правильные ответы. Наличие общепринятой парадигмы, сформулированной в общепринятой системе понятий, означает, что «как думать» – понятно.

Но в науке о мозге до этого еще далеко. А в поисках того, как правильно думать, очень удобным методологическим средством оказывается концептуальная метафора, создающая словарь и круг образов, позволяющий строить гипотезы. Известный пример: атом как солнечная система. Это средство – вспомогательное; его можно отбросить, когда возникнет общепринятая парадигма. Но пока ее нет, концептуальная метафора создает образное пространство для ее поиска, причем ее образы могут подсказать направление поиска – до тех пор, пока не сложится система понятий, которая подходит на роль парадигмы.

Наконец-то когнитивные науки начинают взаимодействовать. Взаимодействие когнитивных наук выглядит примерно так. Нейрофизиология дает сведения о нейронных структурах и их функционировании. При этом не исключено, что нейронные структуры придется, в свою очередь, рассматривать на трех уровнях: сетевом, биохимическом и генетическом. Психология описывает макроэффекты, феноменологические эффекты – поведение, рассуждения, речь – реализация которых должна быть объяснена на нейронном уровне. Искусственный интеллект дает язык, на котором задаются вопросы, систему понятий информатики и средства формализации для построения когнитивных моделей. Когнитивные модели искусственного интеллекта не обязаны утверждать, что мозг работает именно так. Они претендуют лишь на то, что если бы мозг работал согласно модели, то он проявлял бы все те эффекты, которые он проявляет в действительности. Иначе говоря, – эти модели могут претендовать лишь на статус гипотез. Которые должны проверяться науками о мозге.

Для моделирования информационных процессов мозга сетевого языка может оказаться недостаточно. Во всяком случае, ни одна из когнитивных наук в одиночку не способна реализовать проект моделирования информационных процессов мозга. Это можно сделать только вместе.

ЛИТЕРАТУРА

1. *Turing A. M.* Computing machinery and intelligence. *Mind*, v.59. Русский перевод: А. Тьюринг. Может ли машина мыслить? / М., Физматгиз, 1960.
2. *Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И.* Инженерия знаний. Модели и методы / СПб.: Лань, 2016.
3. *Чень Ч., Ли Р.* Математическая логика и автоматическое доказательство теорем. / М.: Наука, 1983.
4. *Вагин В. Н. и др.* Достоверный и правдоподобный вывод в интеллектуальных системах / М.: Физматлит, 2004. – 704 с.
5. *Финн В. К.* Об обобщенном ДСМ-методе автоматического порождения гипотез // Семиотика и информатика. М.: Наука, 1989. Вып.29. С.93-123.
6. *Sowa J. F.* Conceptual Structures – Information Processing in Mind and Machines. Addison-Wesley Publ. Comp. 1984.
7. *Lakoff J.* Women, Fire, and Dangerous Things: What Categories Reveal About the Mind. University of Chicago Press, 1987. Русский перевод: Лакофф Д. Женщины, огонь и опасные вещи: Что категории языка говорят нам о мышлении, М. – 2004.
8. *Hawkins Jeff.* On Intelligence. – New York, Henry Holt and Company, 2005. Русский перевод: Хокинс Джефф. Об интеллекте. – М.: Вильямс, 2007.
9. *Анохин К. В.* Когнитом: гиперсетевая модель мозга // XVII Всеросс. науч.-технич. конф. «Нейроинформатика-2015»: Сб. науч. тр. Ч. 1. М.: НИЯУ МИФИ, 2015.
10. *Анохин К. В.* Когнитивные вычисления на основе нейронных гиперсетей. // Настоящий сборник, с.72-86.
11. *Bullmore E., Sporns O.* Complex brain networks: graph theoretical analysis of structural and functional systems // *Neurosci.* March 2009. V.10. P.186-198.
12. *Sporns O.* Networks of the Brain. Cambridge: The MIT Press, 2011, 412 p.
13. *Кузнецов О. П.* Сложные сети и распространение активности // Автоматика и телемеханика, 2015, № 12, с.3-26.

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ФЕНОМЕНОЛОГИЯ: ВЗАИМОДОПОЛНИТЕЛЬНОСТЬ ИССЛЕДОВАТЕЛЬСКИХ СТРАТЕГИЙ

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE AND PHENOMENOLOGY: THE COMPLEMENTARITY OF RESEARCH STRATEGIES

斯米尔诺瓦 «人工智能哲学与现象学：研究战略互补性»

**Смирнова
Наталья Михайловна**

*доктор философских наук, профессор, главный
научный сотрудник, заведующая сектором
философских проблем творчества Института
философии РАН, г. Москва*

Smirnova Natalia M.

*Doctor of Philosophy, Professor, Chief Researcher, Head
of the Sector of Philosophical Problems of Creativity of
the Institute of Philosophy of the Russian Academy of
Sciences, Moscow*



Аннотация. В статье рассмотрены проблемы философии искусственного интеллекта в предметных срезах философии и когнитивных наук. Выявлен эвристический потенциал феноменологии в исследовании ряда проблем философии искусственного интеллекта. Показано, что, несмотря на различие исходных когнитивных презумпций, феноменология способна внести существенный вклад в решение важных проблем философии искусственного интеллекта – онтологии чистой возможности и ментальной каузальности.

Ключевые слова: философия сознания, искусственный интеллект, нейрофилософия, феноменология, жизненный мир человека.

Annotation: Some basic problems of philosophy of artificial intelligence are studied in the framework of philosophy's and cognitive sciences' subject matters. Heuristic power of phenomenology in the study of some problems of philosophy of artificial intelligence has been discovered. It has been shown, that despite of the difference of initial cognitive assumptions, phenomenology is able to contribute to the solution of some basic problems of artificial intelligence, such as ontology of pure possibility and mental causality.

Keywords: philosophy of mind, artificial intelligence, neurophilosophy, phenomenology, human lifeworld

1. Проблемы искусственного интеллекта в предметных срезах философии и когнитивных наук

Стремительно развивающаяся техногенная цивилизация приводит к глубокой перестройке экзистенциальных смыслов «жизненного мира человека» (Э. Гуссерль). На наших глазах происходят качественные изменения не только экологического, социального и технического ландшафтов человеческого бытия, но и информационно-коммуникативной среды обитания человека. И этот необратимый процесс затрагивает сами основы человеческого бытия – сферу его разума и телесности.

Подобные культурные мутации жизненного мира человека, несомненно, находят свое отражение и в философии как рефлексии над предельными основаниями культуры (В. С. Степин). Под влиянием когнитивных вызовов современности формируются новые области философского исследования. И это процесс объективный, обусловленный когнитивными интенциями современной философии и когнитивных наук, стремящихся проникнуть в «святая святых» самой сложной системы, когда-либо изобретенной матушкой-природой – человеческого мозга, сложность которого на многие порядки превышает сложность всех искусственных систем, когда-либо созданных человеком.

Философия искусственного интеллекта – одна из наиболее динамично развивающихся отраслей философского знания, аккумулировавшая достижения как самой философии, так и когнитивных наук, на обобщения которых она опирается. Ибо в философии искусственного интеллекта, как, быть может, ни в одной другой области философского знания, необходимо «встроить» полученные когнитивными науками данные в общую картину жизненного мира

человека, что, в свою очередь, требует методологического сопровождения философии.

Так, когнитивные науки, на которые опирается философия искусственного интеллекта, предполагают (зачастую неявно) в качестве существенного компонента своих рассуждений идею ментальных репрезентаций. Они мыслятся носителями и хранителями информации, на базе которых и осуществляются вычислительные операции, отождествляемые с мышлением. Ментальные репрезентации обеспечивают контакт познающего существа с миром и определяют его действия. Ментальные репрезентации – современный аналог априорных когнитивных структур на манер идеи внутреннего языка мысли Дж. Фодора, предикаты которого, по его мнению, также являются врожденными. Но исследования В.А. Лекторского свидетельствуют, что идея ментальных репрезентаций, предложенная в качестве научной гипотезы, имеет целый ряд особенностей философского характера [1, С. 6-8]. Это гипотеза высшей степени абстрактности, которую нельзя просто подтвердить или опровергнуть ссылками на отдельные эмпирические факты. Ибо она лежит в основании научной картины мира исследуемой реальности и требует когнитивного согласования отдельных научных положений с этим постулатом.

Другой пример. Нейронауке, например, доподлинно известно, что речевые функции локализованы в так называемой «зоне Брока». При ее повреждении человек теряет способность членораздельной речи. Но это (когнитивно-научное) знание само по себе не дает нам понимания того, как возникает и осуществляется речевая деятельность. Ибо этот факт еще «не встроен» в общекультурную картину человеческого бытия. А в рамках этой картины человеческая членораздельная речь – не только нейрофизиологическая, но и социальная деятельность. И ее исследование возможно на пути продуктивного сотрудничества нейронауки и философского исследования форм коммуникативных практик, т.е. «встраивания» этого открытия нейронауки в общую картину социогенезиса человека.

Но помимо встраивания отдельных научных достижений в общекультурную картину жизненного мира человека, философия выполняет и другую, не менее важную методологическую функцию. Когнитивной ситуации на переднем крае «прорывных» научных направлений, каковыми и являются исследования в области искусственного интеллекта, как правило, свойственен плюрализм подходов и исследовательских стратегий. И философия – и только она

– предоставляет методологические средства анализа когнитивной ситуации конкуренции исследовательских программ в анализе основополагающих проблем искусственного интеллекта. Обозначим лишь две, исходящие из различных когнитивных презумпций и ориентирующие на расходящиеся векторы исследовательских стратегий.

Первая питается когнитивными наработками аналитической философии и прибегает к широкому использованию компьютерной метафоры сознания как вычислительной машины. Адепты этого подхода, например Х. Патнэм, предлагают такое понимание интеллектуальных операций, согласно которому сознание можно уподобить компьютерной программе, осуществляющей вычислительные операции. Впрочем, когнитивные основания для подобной метафоры сложились задолго до появления компьютеров: Г. Лейбниц, к примеру, полагал, что мышление – это скрытая вычислительная деятельность разума, который не знает, что он вычисляет. С появлением первых компьютеров возникла и когнитивная психология, использовавшая «компьютерную метафору» для построения своих теоретических моделей.

Дж. Фодор, считающийся отцом-основателем современной когнитивной науки о сознании, полагал, что интенциональное понятие знания вполне заменимо безличной «информацией». Человеческий мозг перерабатывает информацию по чисто формальным синтаксическим правилам – безотносительно к ее содержанию. Последнее, по Фодору, означает, что исследование когнитивных механизмов работы сознания не требует учета содержательного аспекта поступающей информации. Иными словами, «язык мысли» не нуждается в семантической компоненте, которая вполне может быть элиминирована. Философски это означает, что представление о том, каков мир вокруг нас и что, помимо нашего мозга, представляем собой мы сами, не имеет никакого отношения к анализу когнитивных механизмов работы сознания.

Философские метаморфозы компьютерной метафоры сознания в аналитической философии вылились в продолжительные дискуссии, вышедшие далеко за пределы этого направления философии. В мою задачу не входит обсуждать все ее детали. Ограничусь лишь замечанием, что современная философия сознания вполне осознала ее когнитивные пределы. «Открывая сознание заново», Дж. Серль, по его собственному признанию, вбил последний гвоздь в крышку гроба, в котором упокоилась компьютерная метафора сознания. Сегодня

мы не должны оставаться в плену устаревшей идеи, что одно лишь обладание огромной вычислительной мощью позволяет уподобить искусственный интеллект естественному.

Противоположная исследовательская стратегия исходит из когнитивной презумпции телесной обусловленности человеческого мышления [2]. Она основана на когнитивном допущении, что «не мозгом единым» мыслит человек. Естественный интеллект человека далеко не в последнюю очередь обусловлен его телесной конституцией, сложившейся в процессе генно-культурной коэволюции: в процессах мышления человека участвуют не только его головной мозг, но манер головы проф. Доуэля в фантастическом романе А. Беляева (американский аналог – «мозги в бочке»), но также и нервная, и иммунная системы. Это означает, что ментальность неразрывно связана с телесностью. Имей мы иную генно-культурную историю человечества, то, возможно, не только распределение когнитивных функций в структурах нашего мозга, но и способы нашей презентации в культуре были бы иными.

Продолжая позитивные наработки деятельностного подхода в эпистемологии и философии сознания, телесно-ориентированный подход ориентирует исследователя на изучение энактизации познающих существ: способов его взаимодействия со средой и друг с другом. Как бы то ни было, телесно-ориентированный подход вполне четко артикулирует когнитивный предел уподобления человеческого интеллекта вычислительному устройству.

Попытки философского осмысления ситуации конкурирующих исследовательских программ требуют осознания не только фундаментальных различий их онтологических допущений, но и становления мета-языка обсуждения этой ситуации как философско-методологической, причем такого, который был бы понятен представителям обеих программ. В ответ на эту потребность сформировалась и дисциплинарно оформилась новая область философско-методологических исследований – нейрофилософия [3]. Недавно она обрела и свой периодический печатный орган в России [4], быстро набирающий популярность в научном и философском сообществе.

Следует, по-видимому, согласиться с мнением В. Г. Кузнецова в том, что нейрофилософия ни частично, ни полностью не может считаться общим разделом какой-либо когнитивной науки, и нейронауки, в частности. Она представляет собою самостоятельный раздел философии [5, С. 66]. Сегодня именно нейрофилософия находится

на острие когнитивного интереса современной философии сознания вообще и философии искусственного интеллекта в частности.

Когнитивно-научной основой нейрофилософии являются новейшие достижения нейронауки в исследовании мозговых структур: нейронных сетей, и в особенности так называемых зеркальных нейронов (Rizolatti G., Craighero L.), «отвечающих» за обучение и коммуникацию. Это позволяет, с одной стороны, превзойти представления о «вычислительной» природе естественного интеллекта, с другой – отказаться от абсолютизации селекционистско-адапционистских подходов к пониманию познания. Исследования, проведенные еще в 70-80х годах XIX в., показали, что активность нейрона носит специфический характер в отношении той деятельности, которую выполняет организм. Изучение специфической активности нейронов, а также и более сложных многомерных сетевых структур организации мозговой деятельности – коннектомов и когитомов (Анохин К. В.), – по мнению исследователей, открывает перспективы заглянуть во внутренний мир человека через 10-15 лет.

Справедливости ради отмечу, что налицо и более осторожные прогнозы, основанные на артикуляции фундаментальных различий человека и искусственных технических устройств, имитирующих естественный интеллект. Ибо человека отличает от биоробота наличие субъективной реальности [6], конституированной его субъективным опытом. Корпус субъективного опыта человека представляет собой результат «отложения» в сознании (в терминах социальной феноменологии – седиментации) социальных значений, обретенных в прожитых им «биографически детерминированных ситуациях» (А. Шюц) своей – единственной и неповторимой – жизни.

Проблема теоретического описания субъективного опыта, *qualia*, широко известна под названием «трудная проблема сознания». Суть ее в том, что опыт *qualia* не может быть транслирован в общезначимый язык без потери принципиально значимых смысловых коннотаций. Ибо *qualia* – это опыт «от первого лица». Он принципиально невыразим в общенаучных эталонах мер и весов, т.к. последние не в состоянии описать моего (и только моего) ощущения не только столь тонченных чувств, как любовь, красота, очарование, но даже и относительно простых ощущений тепла, холода, вкуса и боли. В попытках сделать это мы воочию сталкиваемся с семиотическим аспектом «трудной проблемы сознания» – выражения субъективного опыта в интересубъективных терминах. Ибо любое понятие выражает

схематизмы типизированного опыта культурного сообщества, кристаллизованного в общезначимых универсалиях культуры.

Дискуссия о том, как отличить зомби, т.е. существо, имитирующее естественный интеллект, но лишенное субъективного опыта, от человека как социально-исторического существа (тест Тьюринга), имеет давнюю традицию в философии, которая и в настоящее время далека от завершения [7, с. 17-68]. Принципиальное отличие человека от робота, состоящее в том, что «за плечами» естественного интеллекта – огромный пласт культурной истории и личной биографии, не только осмысленной, но пережитой носителем «естественного интеллекта», не может быть элиминировано. Седиментация социальных (интерсубъективных) значений формирует культурный массив личного опыта человека, на основе которого (как правило, не осознавая этого) он совершает поступки и принимает решения в любых жизненных ситуациях. Иными словами, если «поступок» носителя искусственного интеллекта детерминирован программой (пусть и с элементами самообучения), то любой человеческий поступок «дважды детерминирован»: как личной биографией, так и социокультурной историей.

Что же из этого следует? Прежде всего то, что четкая артикуляция когнитивных проблем философии искусственного интеллекта – необходимая предпосылка ее дальнейшего развития. Само наличие таких проблем свидетельствует не о пределах человеческого разума в исследованиях искусственного интеллекта, а о том, что каждый шаг на пути подобных исследований расширяет когнитивный горизонт непознанного – шаг, сделав который мы все более осознаем сложность искусственного интеллекта как предмета исследования. Осознать эти сложности философски – значит задать их культурный масштаб. Но и философия в состоянии сделать это лишь в тесном сотрудничестве с когнитивными науками.

Но о какой философии, собственно говоря, идет речь? Современный историко-философский процесс очень многогранен. Он характеризуется плюрализмом различных концепций и подходов. Т. В. Черниговская, например, особо подчеркивает роль аналитической философии в исследованиях искусственного интеллекта [8, с. 61]. Я же хочу расширить рамки «философского участия» в исследовании проблем искусственного интеллекта и прояснить эвристическую и методологическую роль феноменологии в подобных исследованиях.

2. Философия искусственного интеллекта и феноменология

Методологические наработки философии искусственного интеллекта могут рассматриваться как средства целенаправленного исследования конститутивных механизмов работы сознания. Фантастически расширяя возможности «естественного» интеллекта, они по-новому высвечивают проблемы «предельных оснований» познания. И хотя проблема конституирования (интенциональной) предметности в феноменологии имеет почти вековую историю, конститутивные механизмы работы сознания философски изучены пока недостаточно.

Неявной когнитивной презумпцией классической философии сознания была идея абсолютной «прозрачности» собственного сознания для мыслящего субъекта – полнота саморефлексии как в отношении содержания мышления, так и его предпосылок. Неклассическая философия сознания исходит из фундаментального допущения недостижимости полноты саморефлексии, что, по справедливому замечанию П. Рикера, сначала феноменология, а затем герменевтика непрестанно относили ко все более отдаленному горизонту. Постулат «многослойности» сознания и «непрозрачности» его глубинных слоев для обыденного мышления лежит в основе теории и практики психоанализа – как классического (фрейдовского), так и философски более изощренных его позднейших версий (Фромм, Бинсвангер, Босс и др.).

Важнейшим допущением (пост)неклассической теории познания является представление об изначальной нагруженности «объективного» знания операционально-инструментальными характеристиками деятельности с одной стороны, а также ценностями и жизненными смыслами с другой. (Пост)неклассическая философия сознания сознает культурно-историческую ограниченность классически-рационалистических представлений о сознании как «зеркале души» и разрабатывает философски более изощренные концепции сознания. Крупнейшую брешь в классической философии сознания пробила установка трансцендентализма. «Расщепив» сознание на эмпирическую и трансцендентальную составляющие, И. Кант сосредоточил внимание на исследовании конститутивных механизмов «родового» сознания, т.е. сознания «чистого» (трансцендентального) субъекта. Трансцендентализм Канта сыграл решающую роль в формировании предпосылок «феноменологического поворота» в философии XX столетия.

Но насколько совместимы представления феноменологии о сознании с современными когнитивно-научными исследованиями? Первое, на что я в этой связи хотела бы обратить внимание, – это акцентировать важность тезиса, оказавшегося на периферии философского сознания в парадигме натуралистического поворота в современной философии. А именно: радикальная («сильная») версия натурализма в эпистемологии, заряженная стремлением подменить философию вообще и эпистемологию в частности наиболее общими теоретическими разделами современных когнитивных наук, например, психологии, на практике означает не только девальвацию роли философии в универсуме культуры, но и отказ от решения вопросов, которые, по мысли Э. Гуссерля, «претендуют на высшее человеческое достоинство». И в этой связи аргументы Э. Гуссерля, высказанные им в ходе критики гносеологического психологизма и наивного идеализма в логике и философии арифметики, в целом сохраняют свое значение и для философии искусственного интеллекта.

Подчеркну еще раз, что речь идет именно о содружестве, а не о подмене философии наиболее абстрактными разделами современных когнитивных наук (психологии, нейронауки и т.п.) в духе радикальных версий натуралистического поворота в современной философии [9]. Но и нейрофилософия как современный раздел философского знания [10, с. 65-74], не ставит и не решает специфически-когнитивных проблем нейронауки.

Философия исторически складывалась и продолжает оставаться местоблюстителем высших форм когнитивных абстракций в самосознании культуры, или, по выражению В. С. Степина, «рефлексией над предельными основаниями культуры» [11, с. 190-198]. И это, в первую очередь, относится к феноменологии, о которой далее пойдет речь, – местоблюстителе и хранителе культурной ценности высших форм идеальной предметности в науке и философии. В этом – одна из важнейших функций феноменологии не только в философии искусственного интеллекта, но и в культуре в целом.

Подчеркну еще раз, что именно в свете интенций натуралистического поворота в современной эпистемологии необходимо заново переосмыслить роль философии в современной культуре [12, с. 167-201]. Ибо переход от одной культурной формации к другой неизбежно чреват переосмыслением этой роли. В философии как рефлексии над предельными основаниями культуры кристаллизуется

исторически подвижная матрица, очерчивающая конфигурацию различных предметных сфер человеческого духа.

Почему феноменологическая философия принципиально не сводима к наиболее общим, абстрактным разделам современных когнитивных наук? Если следовать определению М. Хайдеггера, что «мышление есть вопрошание бытия», то феноменология и когнитивные науки ставят по отношению к бытию существенно различные вопросы. Как вариант когнитивного трансцендентализма феноменология обращает к бытию не традиционный вопрос «Что это такое?» («какова его сущность?» и т. п.), но «Как это возможно?».

Нетрудно видеть, что это существенно разные вопросы, обращенные к бытию, и ориентируют они на существенно различные исследовательские стратегии. В отличие от когнитивных наук, феноменологический анализ разворачивает исследование в противоположном направлении. Феноменология устремлена к выявлению и элиминации скрытых (а потому и далеко не очевидных) предпосылок (особенно метафизического характера), обусловивших саму возможность того или иного явления в его чистых логических формообразованиях. Не слишком огрубляя когнитивные интенции обеих стратегий, их можно было бы соотнести как разнонаправленные (но взаимодополнительные) векторы когнитивного интереса.

В когнитивных науках исследовательская мысль движется от явления к его сущности, в феноменологии – от сущности к «чистому» явлению. Абстрагируясь от онтологических и иных предпосылок естественной установки сознания, трансцендентально-феноменологическая редукция выводит исследователя на уровень чистой беспредпосылочности и эксплицирует условия образования идеальных предметностей как принципиальную возможность для Разума.

Известно, что еще в 18 в., т.е. в период «классического» развития естествознания (используя периодизацию науки В. С. Степина), в естествознание проникает идея эволюции. Философы также воздают дань этим переменам в науке – вспомним «докритический» (до 1770 г.) период научной деятельности Канта, воплощенный в его труде «Всеобщая естественная история неба». С тех пор когнитивные науки нацелены на познание своего предмета в его становлении и развитии. И не в последнюю очередь потому, что свойственная классической науке эссенциалистская установка ориентирует исследователя на поиск неизменного, инвариантов различных стадий развития

предмета, что позволяет выявить его сущность как неизменное при всех изменениях.

Когнитивные науки изучают изменение бытийных характеристик предмета как результат его развития во времени, прибегая к сбору эмпирических данных путем наблюдения, сравнения, экспериментов. Феноменологии, которая, как и философия в целом, является, по Э. Гуссерлю, «строгой наукой», также присущи интенции «усмотрения сущности как неизменного во всех отношениях». Однако для этой цели она прибегает не к идентификации отдельных стадий развития предмета в реальной истории его становления. Ей присущ особый метод поиска сущности – феноменологический метод «интуитивного усмотрения» (*Wesensschau*). Феноменологический метод выявления сущности состоит в мысленном варьировании предмета с целью выявления его инвариантных свойств. Например, сущность «кубичности» мы можем уяснить, мысленно «вращая» рассматриваемый предмет и «вычитая» (в терминологии Э. Гуссерля – «заклучая в скобки», «выводя из игры») его неатрибутивные свойства. В случае поиска сущности «кубичности» мы можем феноменологически «заклучить в скобки» его размер, цвет, плотность, материал, из которого он сделан, текстуру поверхности и т.п. Но мы, очевидно, не сможем посягнуть на равенство его ребер или прямых углов наклона его боковых поверхностей к основанию, если хотим, чтобы предмет оставался кубом. Эти характеристики предмета и будут в данном случае сущностными характеристиками «кубичности». Таким образом, хотя методы поиска сущностных характеристик у феноменологии и когнитивных наук различны, тем не менее, как когнитивные науки, так и феноменология устремлены к одной цели: поиску сущности как неизменного (относительно устойчивого) во всех проявлениях. Именно поэтому мы вправе говорить не только о противопоставлении, но и взаимодополнительности методов феноменологии и когнитивных наук на основе их общих когнитивных устремлений.

Абстрагируясь от бытийных («посюсторонних») характеристик феномена, данных в «естественной установке сознания», феноменолог осуществляет операцию трансцендентально-феноменологической редукции и выходит на уровень чистых феноменов сознания как центров мировых отношений – чистого сознания трансцендентального Эго. Это аналог виртуальной реальности современных компьютерных технологий. Феноменология отвлекается от онтологического статуса идеальных предметностей – ее интересует лишь

его логическая возможность как в принципе доступная для Разума. Можно было бы сказать, что феноменологически «существовать – значит быть возможным», если бы в трансцендентально-феноменологической сфере различие между возможным и действительным сохраняло свое привычное для естественной установки сознания значение.

Полагаю, что к философии искусственного интеллекта вполне применим аналогичный трансцендентально-феноменологическому принцип полноты реализации возможного, развитый в рамках модального реализма Д. Льюиса. Согласно этому принципу, никакая подлинная потенция бытия не может не быть реализована: все возможные миры реальны. Трансцендентально-феноменологическая онтология, равно как и онтология философии искусственного интеллекта – это онтология чистой возможности, тождественной действительности, отвлеченной от вопросов посясторонней бытийности – абсолютное сознание, неотличимое от его содержания. И в этом ее принципиальное сходство с виртуальной реальностью, моделируемой современными ИТ – технологиями.

Так какое же отношение она имеет к реальности, изучаемой когнитивными науками? Дело в том, что в трансцендентально-редуцированной сфере все базовые значения естественной установки сознания сохраняются, но предстают в более отчетливом, трансцендентально-чистом виде – сам Э. Гуссерль считал этот факт одним из наиболее удивительных, но в то же время – принципиально значимых. Поэтому наработки феноменологов, работающих в сфере онтологии чистой возможности при отвлечении от бытийных вопросов существования, обладают мощным эвристическим потенциалом для компьютерного моделирования в такой сравнительно молодой области когнитивных исследований, как компьютерный инженеринг, нейрокомпьютинг и – шире – философия искусственного интеллекта.

И еще один важный момент в плане методологии. На примере феноменологии мы можем видеть, что задачей философии является вовсе не поставка готовых методов когнитивным наукам. Ее задача, скорее, – методологическое сопровождение и выявление скрытых трудностей в процессе когнитивно-научного исследования, обусловленных традиционным невниманием к неявным предпосылкам и привычкой изучать теоретические вопросы в отрыве от методологических. Но с другой стороны, основатель феноменологии Э. Гуссерль полагал, что когнитивная наука не является и не может

быть окончательной познавательной критикой разума: механизмов внутринаучной саморефлексии, по его мнению, недостаточно для отыскания подлинных оснований научной достоверности – «твердой породы» научного, философского, да и любого рационального суждения вообще. Сама наука нуждается в феноменологическом прояснении ее глубинных когнитивных оснований посредством феноменологической критики.

В условиях господства позитивизма в европейской философии начала XX в., укоренившегося в методологическом самосознании науки, Э. Гуссерлю пришлось отстаивать подобные взгляды в острой полемике не только с представителями когнитивных наук, но и с коллегами-философами позитивистской ориентации. И дальнейшее развертывание его феноменологической эпистемологии призвано обосновать общекультурную роль феноменологии в поддержании рационального тона европейской культуры в условиях грядущего «кризиса европейских наук».

Однако «хитрость мирового разума» состояла в том, что когнитивными науками оказались востребованы не совсем те идеи, которые сам Э. Гуссерль считал наиболее существенными, во всяком случае, в ранний период своего философского творчества, сфокусированного на проблемах эпистемологии.

Исторически сложилось так, что первая когорта последователей Э. Гуссерля, близко знакомых с ним лично (Эдит Штайн, Герда Уолтер и отчасти А. Шюц), поначалу полагали, что конкретные проблемы когнитивных наук можно разрешить с помощью непосредственного приложения к ним феноменологических методов, в частности метода трансцендентально-феноменологической редукции. Они полагали, что именно с помощью этого метода возможно феноменологическое прояснение базовых понятий когнитивных наук в целях достижения их картезианской «ясности и отчетливости». Сходной же точки зрения поначалу придерживался и один из наиболее выдающихся учеников Э. Гуссерля М. Шелер. Однако попытки непосредственного («прямого») приложения феноменологических методов без надлежащей «настройки» на предметную специфику различных когнитивных дисциплин привели к обратному результату. Идеи феноменологии в глазах социальных ученых предстали чересчур абстрактными и весьма далекими от конкретных методологических проблем и когнитивных запросов отдельных научных дисциплин.

Однако идеи трансцендентальной феноменологии способны эффективно воздействовать на когнитивные науки на уровне их философских оснований (концепт В.С. Степина) и лишь частично преломлялись в их исследовательской «технике». Так, центральное место в феноменологическом подходе к анализу артефактов культуры занимает разработанное в поздних работах Э. Гуссерля понятие «жизненный мир человека» (*Lebenswelt*). Жизненный мир – это совокупность дорефлексивных очевидностей, данных в «естественной установке» сознания. Понятие жизненного мира манифестирует возврат из сферы чистого сознания трансцендентального Эго в область повседневности и социальной жизни. В соответствии с общей установкой феноменологии жизненного мира на исследование смысловой структуры социального мира, жизненный мир человека в социально-философском контексте предстает как квинтэссенция опыта intersубъективного социального отношения.

Восходя к методологической рефлексии современного Э. Гуссерлю естествознания, эвристический потенциал понятия жизненного мира имеет общекультурную размерность. В последней из опубликованных при жизни основателя феноменологии работе – «Кризис европейских наук и трансцендентальная феноменология» – он неоднократно подчеркивал, что любые используемые человеком понятия, включая и самые изощренные научные абстракции, имеют глубокие неотрефлексированные корни в жизненном мире человека [13]. Они обнаруживают себя в естественном языке, на котором основан язык науки (и эта зависимость неустранима в силу невозможности чисто «объектного» языка), а также мировоззренческих презумпциях, не проблематизируемых («само собой разумеющихся») в рамках определенных универсалий культуры. Поэтому одной из важнейших задач методологической саморефлексии науки Э. Гуссерль полагал исследование корней специализированного научного знания в жизненном мире человека – процесса седиментации научных значений. По Э. Гуссерлю, любые научные понятия являются не более чем изощренной формой интеллектуальной стенографии, если не выявлены их глубокие жизнемировые корни. Ибо вырастают они на почве жизненного мира и коренятся в структурах естественной установки сознания.

Методологически апелляция к жизненному миру изменяет место теоретической речи. Говорящий от имени науки утрачивает статус абсолютного наблюдателя, вознесенного над жизнью, и предстает как

изначально погруженный в структуры жизненных миров. Презумпция изначальной «захваченности» социального теоретика структурами жизненных миров методологически означает, что он принципиально не свободен от предпосылок, разделяемых с обыденным мышлением в естественной установке сознания. Предпосылки такого рода являются глубокими отложениями реального жизненного опыта и, как правило, носят неявный характер, подлежащий феноменологическому «расколдованию».

Анализируя истоки и смысл кризиса научной рациональности как манифестации общего кризиса европейских наук, Э. Гуссерль показывает, что понятие природы в классической механике является ничем иным, как идеализированной абстракцией жизненного мира, – абстракцией, элиминирующей человека и артефакты культуры. «Включала ли она (реальность математизированной науки – Н. С.) человека? – вопрошает последователь Э. Гуссерля Т. Лукман. – Декарт включил в этот мир человеческое тело, но, увы, не душу. Единство старого мировоззрения раскололось: радикально различные стили мышления предписывались знанию о человеке (и Боге) и остальному универсуму» [14, с. 226-227].

Как показывают исследования исторических форм научной рациональности В. С. Степина, реифицирующая установка приводит к тому, что классическое естествознание обнаруживает в природе слепки собственной теоретической и инструментальной деятельности. Онтологизация научных процедур элиминирует проблему относительности научного результата к операциям и средствам деятельности, что приводит к подстановке небеспредпосылочных и теоретически нагруженных научных конструкторов на место характеристик самой природы. Так происходит замещение реальности жизненного мира искусственной реальностью идеализированных объектов науки, а ученый принимает за характеристики природы продукты собственной объективирующей методологии.

И еще один методологически важный вопрос. Гуссерлево понятие жизненного мира высоко эвристично в решении вопроса о соотношении ментальной каузальности и физической причинности, широко обсуждаемого в настоящее время, в том числе и в рамках философии искусственного интеллекта. Вопрос этот имеет давнюю историю. Так, например, Р. Пенроуз [15], продолжая традиции натуралистической тенденции в эпистемологии, предлагает два принципиальных решения, каждое из которых располагается (хотя и

по-разному) в рамках натуралистической парадигмы в философии. Первый – так называемая квантовая теория сознания, согласно которой объяснение «теньям сознания» (*shadows of the mind*) следует искать в трансформациях на квантовом уровне, возникающих в основаниях нейронов головного мозга. Квантовая теория сознания активно обсуждается в нейрофилософии, однако, как мне представляется, судить о ее когнитивных достоинствах сложно, пока в рамках самой нейронауки не получены философски значимые выводы.

Второй вариант предполагает существенную трансформацию всей физической картины мира. Но и этот вариант не представляется перспективным. Ибо физическая картина мира – методологически сложный конструкт, исторически сложившийся в научном сообществе и не поддающийся произвольному изменению. Научная картина мира сложным образом взаимодействует как с онтологическими проекциями отдельных научных теорий, так и с мировоззренческими универсалиями культуры и не может изменяться произвольно, повинувшись когнитивным интенциям *ad hoc*- гипотез. Поэтому наиболее эвристичным представляется третий путь, «твердой породой» (когнитивным фундаментом) которого является богатство содержаний, заложенных Э. Гуссерлем в понятии «жизненный мир человека».

Мой тезис состоит в том, что концепт жизненного мира помогает избежать дуализма этих видов причинности. Если, следуя Э. Гуссерлю, признать тезис первичности жизненного мира человека, то физический мир предстанет его идеализированной (или, если угодно, «тощей») абстракцией. Приняв же традиционную, сугубо натуралистическую установку, полагающую онтологическую первичность физического мира, отвлеченного от человека и его культурных артефактов, нам никоим образом не удастся избежать методологически некомфортного дуализма причинности. Кроме того, подобная «жизнемировая» установка, как нетрудно видеть, вносит весомый вклад в решение проблемы онтологического статуса артефактов. В самом деле, обладают ли камень на дороге и мраморная статуя одинаковым бытийным статусом? С позиции заложенной Э. Гуссерлем жизнемировой установки, безусловно, да.

Понятие жизненного мира Э. Гуссерля позволяет переосмыслить привычное соотношение реальности и нашего места в ней. Мы привыкли считать само собой разумеющейся онтологическую первичность природы, тогда как «очеловеченный» мир, светящийся смыслом, уютно укладывается в наличные культурно-исторические

схематизмы нашего мышления, деятельности и социальной организации. Иными словами, в наших классически-объективистских (дофеноменологических) представлениях мир человека привычно «надстраивается» над миром природы.

Фундаментализм жизненного мира кардинально меняет привычную оптику наивного онтологизма «природа – сознание – культура». Мир человека, универсум овеществленных артефактов культуры и социальности полагается первичным как фундаментальная культурно-антропологическая реальность, изначально данная человеку. Мир же природы – идеализированная абстракция жизненного мира, результат искусственного «заклучения в скобки», «выведения из игры» (Э. Гуссерль) философски самого существенного: человека с его обжитым миром культуры и социальности. В мире, где есть человек и очеловеченная им природа, мир «сам по себе», «природа до и вне человека» – сильнейшая идеализация абстрагирующего мышления Нового времени, лежащая в основании классического естествознания. Фундаментализм жизненного мира, т.о., философски «снимает» как наивный онтологизм классически понятой объективности, так и дуализм физической и ментальной каузальности. В философской разработке онтологии чистой возможности, а также в устранении дуализма ментальной каузальности и физической причинности я и усматриваю важнейшую роль феноменологии в ее отношении к когнитивным наукам.

ЛИТЕРАТУРА

1. *Лекторский В. А.* Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы // Естественный и искусственный интеллект / Отв. ред. Дубровский Д. И., Лекторский В. А. – М., Канон+ РООИ «Реабилитация», 2011. – 352 с.
2. *Князева Е. Н.* Энактивизм: новая форма конструктивизма в эпистемологии. М., СПб, Центр гуманитарных инициатив; Университетская книга, 2014. – 352 с.
3. Основателями нейрофилософии, ведущей отсчет с 60-х г.г. XIX в., считаются Пол и Патрисия Черчленд (США): *Churchland P. Neurophilosophy.* Cambridge, MA: MIT Press, 1986.
4. *Актуальные вопросы нейрофилософии.* Серия «Нейрофилософия». Ежегодник /Отв. ред. Алексеев А. Ю. и др.. М., ИИнтелл, 2015. – 284 с.
5. *Кузнецов В. Г.* Нейрофилософия и обоснование внутреннего мира человека // Актуальные вопросы нейрофилософии. Серия «Нейрофилософия». Ежегодник /Отв. ред. Алексеев А. Ю. и др.. М., ИИнтелл, 2015. – 284 с.
6. *Дубровский Д. И.* Проблема «сознание и мозг»: теоретическое решение. М., Канон+. 2015. – 208 с.

7. *Алексеев А. Ю.* Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИИнтелЛ, 2013. – 304 с.
8. *Черниговская Т. В.* Конь и трепетная лань: ученый на стыке наук // Актуальные вопросы нейрофилософии. Серия «Нейрофилософия». Ежегодник /Отв. ред. Алексеев А. Ю. и др.. М., ИИнтелл, 2015. – 284 с.
9. Так, наиболее заметный представитель натуралистического подхода в современной эпистемологии П. Черчленд, считающаяся основателем нейрофилософии, полагает, что нейробиологические данные имеют прямое отношение к давним, традиционным аспектам философии, таким, как природа сознания, самость («Я»), свобода воли, знание, этика, научение.
10. *Алексеев А. Ю., Кузнецов В. Г., Савельев А. В., Янковская Е. А.* Становление отечественной нейрофилософии // Философские науки, 2015, № 11; *Кузнецов В. Г.* Нейрофилософия и обоснование внутреннего мира человека. Актуальные вопросы нейрофилософии. Серия «Нейрофилософия». Ежегодник /Отв. ред. Алексеев А. Ю. и др.. М., ИИнтелл, 2015.
11. *Степин В. С.* Философская антропология и философия культуры. – М., Академический проект, Альма Матер, 2015. – 542 с.
12. См: *Смирнова Н. М.* Философия, специализированное социальное знание и жизненный мир человека // Наука глазами гуманитария /Отв. ред. Лекторский В. А. М., Прогресс-Традиция, 2005..
13. См: *Husserl E.* Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Hamburg: Nijhoff, 1977. Русск. Пер.: Гуссерль Э. Кризис европейских наук и трансцендентальная феноменология. СПб, Наука, 2013.
14. *Luckman Th.* Philosophy, Social Sciences and Everyday Life // Phenomenology and Sociology. N.Y.: Penguin Books, 1978. P. 226-227.
15. *Penrose R.* Shadows of the mind: A Search for the missing science of consciousness. Oxford univ. press, Oxford, 1994 – 457 p.

ТЕСТ ТЬЮРИНГА: 66 ЛЕТ СПУСТЯ TURING TEST: 66 YEARS LATER

阿列写扶, 余 «图灵(«TURING»)试验 : 66
年后»

Алексеев

Андрей Юрьевич

*координатор научных программ НСММИ РАН,
доктор философских наук, ведущий научный
сотрудник кафедры философии и методологии
науки философского факультета МГУ имени
М. В. Ломоносова, г. Москва*

Alekseev Andrey Y.

*Coordinator of scientific programs of the RAS
SCMAI, Doctor of Philosophy, leading researcher of
the Department of Philosophy and Methodology of
Science, Faculty of Philosophy, Lomonosov Moscow
State University, Moscow*



Аннотация. Определяется философское содержание идеи игры в имитацию А. Тьюринга (1950 г.) Изучается история становления комплексного теста Тьюринга. Обозначаются функции этого теста в анализе проблем электронной культуры: интеррогативная, дефинитная, критическая, конструктивная, конститутивная.

Ключевые слова. А. Тьюринг, игра в имитацию, комплексный тест Тьюринга, машинный функционализм, тестовый компьютеризм

Annotation. The philosophical content of the idea of the game in the imitation of A. Turing is determined (1950). The history of the formation of the Comprehensive Turing test is being studied. The functions of this test are identified in the analysis of problems of electronic culture: interrogative, definitive, critical, constructive, constitutive.

Keywords. A. Turing, game in imitation, complex Turing test, machine functionalism, test computationalism.

В этом году искусственный интеллект отмечает две юбилейные даты. Во-первых, шестидесятилетие научного направления исследований: этому празднику посвящено сегодняшнее мероприятие. Другая интересная дата, без которой не было бы и первой: 66 лет назад возникла сама идея построения интеллектуальных компьютеров, точнее, цифровых электронных вычислительных машин, некоторым функциям которых некоторые люди приписывают некоторые характеристики человеческого разума. А именно, в 1950 году в престижном журнале «Mind» была опубликована небольшая статья Алана Тьюринга [1], которая сегодня стала самой известной в когнитивном и компьютерном мире. На русский язык статья была переведена в 1960 году под редакцией Б. В. Бирюкова как «Вычислительные машины и интеллект» [2]. На нашей конференции о роли этой статьи для искусственного интеллекта и современной философии немного уже упоминал Лекторский Владислав Александрович (см. настоящий Сборник). В своем докладе я более детально проанализирую историю развития тьюринговой идеи, обрисую ее текущее состояние и оценю вклад в развитие искусственного интеллекта.

О тесте Тьюринга мне посчастливилось говорить ранее в авторитетной аудитории: на 55-, 60- и 65-летних рубежах истории развития теста Тьюринга [3; 4, с. 8-19; 5]. Лейтмотивом этих выступлений была идея о том, что современный тест Тьюринга – это уже не тот оригинальный тест, который весьма наивно (конечно, «наивно» по нынешним меркам), был сформулирован А. Тьюрингом в 1950 году. Сегодня тест Тьюринга следует воспринимать с учетом многочисленных, весьма значительных изменений, переформулировок, модификаций, доработок, то есть рассматривать его с учетом версий, версий этих версий, версий версий и т.д. А таких переработанных, часто до неузнаваемости, тестов – более сотни.

Поэтому достойно оценить вклад главной идеи искусственного интеллекта – теста Тьюринга – мы сможем, когда систематически изучим и интегрируем эти «версии версий», выявим, обобщим и доопределим основные функции, которые выполняет такой комплексный тест в современном и перспективном информационном обществе, наметим пути предсказания когнитивных явлений развивающегося компьютерного мира.

План моего доклада таков: 1) небольшой исторический экскурс в становление идеи комплексного теста; 2) определение комплексного теста Тьюринга и выявление его основных функций; 3)

определение сферы применения комплексного теста; 4) обозначение роли теста в когнитивной философии; 5) выводы относительно практических приложений комплексного теста. Многие мои сегодняшние идеи изложены в специальной монографии [7], поэтому буду предельно краток.

История развития идеи теста Тьюринга

Так как отмечается своеобразный 66-летний юбилей теста, то нельзя обойти вниманием историю становления и развития дискуссий по поводу теста Тьюринга. Забегая вперед, отмечу, что сегодня в аналитической философии искусственного интеллекта рубрика «Тест Тьюринга» занимает первое место. По крайней мере, так убеждает известная «Библиография работ по философии сознания и философии когнитивной науки», составленная австралийскими философами-метафизиками Дэвидом Чалмерсом и Дэвидом Бурже (David Bourget). В этой библиографии рубрика «Философия искусственного интеллекта» была открыта темой теста Тьюринга (до 2012 года библиография находилась по адресу <http://consc.net/mindpapers>).

Конечно, в деле тестирования когнитивных компетенций у машин А. Тьюринг не был первым. Задолго до него в XVII веке в классической философской работе «Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках» Рене Декарт обсуждал подобную тему (в V части). Он доказывал, что даже если машина никак не будет отличаться от человека – ни телом, ни поведением, – то все равно можно отличить одного от другой. Для этого надо протестировать машину на её способность пользоваться языком так же, как пользуются люди, сообщая другим свои мысли. Машина провалит тест, так как нельзя представить, что она должным порядком будет располагать слова. И, во-вторых, надо протестировать машину на её способность работать в самых различных контекстах. Тест не будет пройден, так как только человеческому разуму под силу охватить всё многообразие всевозможных поведенческих ситуаций. Хотя в некоторых частных случаях машина может действовать лучше человека. Да, кстати, машина и животное при таких же условиях тестирования будут неразличимы, животные ведь не обладают разумом, как полагает Декарт.

Тьюринг также тестирует машину на предмет языковых компетенций, однако категорически оптимистичен относительно возможностей программируемых машин. Он предсказывает, что через полвека, то есть в 2000 году, лингвистические способности человека и

компьютера станут неразличимыми. То есть, в отличие от Р. Декарта, А. Тьюринг положительно решает вопрос построения компьютеров, способных во всём, включая осмысление предметной области, подражать человеку.

Словосочетание «Тест Тьюринга» сейчас у всех на слуху. Однако собственное наименование этому тесту присвоили не сразу, лишь в конце 1960-х годов. В статье [1] А. Тьюринг описывает не тест, но несколько сценариев игры в имитацию диалога на естественном языке, который изначально ведется для выявления пола игрока: с мужчиной или женщиной беседует судья (наблюдатель)? Слово «тест» не употребляется в смысле тестирования человека и программируемого компьютера на предмет выявления у них интеллектуальных функций. В большей степени интересует правдоподобность компьютерного сценария игры в имитацию. Для этого вводятся этапы игры в имитацию. На каждом из этапов компьютер все больше и больше «захватывает» власть по реализации форм осмысленного диалога.

На первом этапе играют люди – мужчина и женщина. Судья их не видит и общается письменными сообщениями коммуникационными средствами телетайпа, то есть тем, что сегодня называется электронной почтой, чатом, ботом, но не скайпом и тем более не мультимедийными средствами передачи запахов и рукопожатий, – судья не видит собеседника. Женщина говорит только правду, мужчина может вводить судью в заблуждение, прикидываясь женщиной. Конечно, при таких условиях судья только наугад способен предсказать правильный пол: «50 на 50, и зачем это надо?» – примерно так отреагировала Софья Александровна Яновская, автор предисловия к книге «Могут ли машины мыслить?» (1960 г.), создатель советской школы философии математики и организатор кафедры математической логики на механико-математическом факультете МГУ.

Однако логическая корректность сценария не волнует автора статьи. Ведь главное – продемонстрировать способность компьютера имитировать человеческий диалог. Важно то, насколько часто судья будет ошибаться при общении с человеком и компьютером. Условно введенная вероятность ошибки – 30 процентов неправильного определения пола при пяти минутах «общения», хотя и теряет смысл в контексте некорректных начальных условий, тем не менее не существенна для обозначения принципа компьютерной имитации интеллектуальных способностей, проявляемых в условиях повседневного общения людей. Для этого на второй стадии игры в имитацию роль

женщины играет компьютер (здесь и возникают отмеченные выше казусы с вероятностями). А на третьей стадии компьютер уже захватывает роль мужчины, пол игрока становится несущественным, и, собственно, возникает возможность сформулировать то, что принято считать каноническим тестом Тьюринга, точнее, его двумя формами. Имеются две разновидности канонического теста Тьюринга. Они сегодня общеприняты. Первая имеет дистрибутивный, разделительный формат, а вторая – атрибутивный формат, приписывающий человечность или компьютерность тестируемой системе.

Дистрибутивная форма. В игру имитации интеллекта играют и люди и компьютеры. Задача судьи – определить, кто есть кто? Кто из собеседников – компьютер, а кто – человек? Эта форма сейчас используется в Лойбнеровских состязаниях. Об этих «тьюринговых скачках» я упомяну в заключение доклада.

Атрибутивная форма. В игру в имитацию интеллекта играет неизвестный субъект. Задача судьи выявить «человечность» либо «машинность» собеседника.

Это взаимосвязанные формы: определить – значит разграничить и наоборот. И та и другая формы значимы. Например, они непосредственно характеризуют информационную безопасность. Для судьи важно не различие компьютерного/человеческого, а выявление факта общения именно с человеком. С кем я пойду сегодня на свидание? Кто просит меня переслать электронные деньги? Кто мне навязывает политический имидж президента? Это – автономная программа или реальный человек? Но какие фрагменты «человека-компьютера» можно запрограммировать и кто на это способен?

Как отмечалось выше, сам А. Тьюринг в этой статье не употреблял слово «тест» в смысле тестирования компьютера на интеллект. И это послужило поводом для ехидных замечаний: в мире когнитивной философии А. Тьюринг стал известен за то, что, собственно, никогда и не предлагал – он не сформулировал «тест Тьюринга» (Дж. Лассаж, 1996). Однако это не совсем так. На самом деле, А. Тьюринг в 1952 году в репортаже на студии БиБиСи говорит о своем мысленном эксперименте не как об игре в имитацию, но как о тесте. Правда, в этой беседе, указав на сложность задачи моделирования мозга и интересубъективных отношений, А. Тьюринг отодвинул сроки реализации такого теста еще на пятьдесят лет вперед [9, Р. 126]. То есть, по его новому предсказанию, не в 2000 году, а в 2052 году следовало бы ожидать появление «компьютерного собеседника по переписке

от рождения до смерти». Если бы прогноз сдвигался и далее с такой прогрессией динамики «знаний о незнании» (ведь в 1950 году А.Тьюринг не знал о сложности задач, которую оценил спустя два года), то сегодня, в 2016 году, следовало бы ожидать прохождения теста Тьюринга не ранее 3716 года. Это было бы к великому сожалению для устроителей Лойбнеровских состязаний, стремящихся буквально следовать тонкостям тьюринговой идеи игры в имитацию.

Конечно, столь догматичному толкованию игры в имитацию не следуют нормальные исследователи теста Тьюринга. Для них тест служит концептуальным каркасом искусственного интеллекта, является генератором новых идей, различных интерпретаций и толкований. Исследуются самые разные параметры теста: предмет тестирования – интеллект, сознание, переживание и пр.; длина теста – интервал времени, в течение которого осуществляется тестирование; содержание тестовых вопросов; стена Тьюринга – параметры скрытности игроков от судьи; особенности игроков и судей и многое др. На разнообразии трактовок влияет неопределённость идеи игры в имитацию, «чехарда» с ролями мужчины, женщины, машины и пр.

Реакция на статью возникла, впрочем, не сразу: 15- 20 лет дискуссии не выходили за пределы журнала «Mind». Причем вряд ли эти дискуссии можно назвать серьезными. Так, Кейт Гандерсон в статье 1964 г. предлагает поиграть в игру «в наступалки»: судья просовывает ногу в дырку стены комнаты и на его ногу начинают наступать в различном порядке мужчина и женщина. Далее вместо женщины начинает работать машина, которая посредством видеодатчиков запускает механизм падения камней на просунутую ногу. Далее эта машина начинает играть роль мужчины. Получается вопрос «Могут ли камни подражать?». И что из этого следует? Другие «исследователи» вместо ленты Тьюринга предлагают использовать туалетную бумагу и поднимают вопрос: какой длины должна быть бумага, чтобы возникла иллюзия имитации интеллекта? А Ричард Пёртилл в 1971 г., выступая как продвинутый программист, но не как философ, предлагает съесть электронную библиотеку у того, кто спроектирует компьютер, успешно играющий в игру в имитацию. Игра в имитацию – это из сферы научной фантастики, считает он. Однако в этом же году К. Колби в работе «Искусственная паранойя» (1971 г.) представил описание обильно профинансированной реализации игры в имитацию компьютером ответов пациентов психиатрической лечебницы.

Врачи – судьи затруднялись ответить, от людей или компьютеров поступают ответы.

Некоторые исследователи занялись плагиатом тьюринговой идеи. Так, Джон Сёрль в 1980 году в своем знаменитом мысленном эксперименте «Китайская комната» почти досконально воспроизвел тьюринговский сценарий, лишь немного его изменив: роль человека и компьютера играет сам Дж. Сёрль, снабженный компьютером с базой данных-иероглифов, а роль судьи попеременно играют китаец и англичанин. Ну не его этот сюжет: Дж. Сёрль на самом деле забрел в комнату Тьюринга, но хоть как-то для приличия надо бы было обозначить ссылку на сюжет тридцатилетней давности.

Многие исследователи, однако, отнеслись серьезно к игре в имитацию. Например, Нэд Блок в 1981 году предлагает т.н. «новый тест Тьюринга», в котором обращает внимание на важность смысловых параметров диалога, а не только на синтаксис и семантику сообщений. Также он формулирует отличие интеллектуальной системы от системы искусственного интеллекта, предвосхищает роль интернета как глобальной интеллектуальной компьютерной системы, в которой, правда, нет ни толики искусственного интеллекта, формулирует концепцию психофункционализма – одного из продуктивных направлений моделирования когнитивных систем и пр.

В Советском Союзе отношение к идее А. Тьюринга с самого начала было весьма почтительным. Во-первых, следует учесть тираж книги «Могут ли машины мыслить?» – 50 тысяч экземпляров! Причем книга мгновенно распространилась по всей стране. Сразу повсеместно разгорелись жаркие дискуссии, появились научно-фантастические фильмы и повести про искусственный интеллект. В 1963 г. академик А. Н. Колмогоров пишет статью «Автоматы и жизнь», где утверждает теоретическую возможность воспроизводства автоматами всех видов человеческой активности и не только интеллектуальных, но и эмоциональных. На это Б. Бялык отвечает статьей «Товарищи, вы это серьезно?». В ответ академик С. Л. Соболев пишет работу «Да, это вполне серьёзно!». Завязалась серьезная полемика, в рамках которой обсуждались вопросы возможности кибернетического определения таких понятий, как воля, мышление, эмоции и др.; противоречит либо соответствует материалистической диалектике концепция «думающих» машин; как понимать «машину» в контексте исторического подхода, ведь машина – это продукт общественно-трудовой деятельности человека, машина не трудится, трудится

человек посредством машины и пр. Возникли этические проблемы контроля над «интеллектуальными» машинами.

Вскоре в нашей стране споры по поводу мыслящих машин, однако, утихли. Как, впрочем, были свернуты интереснейшие проекты советских ЭВМ, таких как трехзначная машина «Сетунь» (разработка МГУ). Примерно в эти же годы протоколы вычислительных сетей, разработанные в отечественных НИИ, оказались в ARPANET под новым названием «TCP/IP». Произошел ряд других событий, непонятных с позиции национального научно-технического прогресса.

За рубежом интерес к тесту Тьюринга ярко вспыхнул в 1990-е годы. Появились крупные статьи. В аналитической философии сознания прочные позиции занял компьютерный функционализм. В 1991 году продавец светящихся полов для дискотек Хьюго Лойбнер (Hugh Loebner) открыл ежегодные тьюринговые игры, впервые предложив 100 тысяч долларов за программу, которая пройдет тест Тьюринга. В общем, тьюринговая гонка началась.

Сегодня насчитываются сотни крупных статей, монографий, диссертационных исследований и специальных работ, посвященных тесту.

Частные тесты Тьюринга

В литературе выделяется определенное количество крупных версий теста Тьюринга. Условно говоря, имеется двадцать одна версия. Эти версии обозначены как *совершенные* частные тесты Тьюринга, так как в них четко просматриваются базовые функции, предложенные в оригинальном тесте: 1) интеррогативная функция, характеризующая содержание и форму вопросов частного теста, которые задает тьюринговый судья (interrogator, наблюдатель) при изучении *x*-системы на предмет ее *y*-способностей; 2) дефинитная функция, обеспечивающая компьютерное определение когнитивного феномена; 3) конструкторская функция, раскрывающая принципы работы компьютера, способного реализовать когнитивный феномен; 4) критическая функция, отражающая суть полемики по поводу возможности компьютерной реализации когнитивного феномена; 5) конститутивная функция, позиционирующая отношение судьи (наблюдателя) к когнитивным аспектам компьютерной реальности.

Помимо совершенных версий, имеется большое количество несовершенных. В них акцентируется внимание, на, возможно, интересной, но незначительной детали тьюрингового тестирования. Таких несовершенных версий много, более сотни зарегистрированных.

Кратко перечислю совершенные частные тесты, которые, на мой взгляд, составляют ядро комплексного теста Тьюринга. Подробно они рассмотрены в книге [7].

1) Оригинальный тест. Отвечает на вопрос: может ли компьютер мыслить? (А. Тьюринг, 1950). Изучает возможности компьютерной имитации осмысленного диалога. Применяет концепцию универсальной УЦВМ и параллельной УЦВМ – так называемой «нейронной сети Тьюринга».

2) Тест на «здравый смысл». Включает тест на глубинный «здравый смысл»: может ли компьютер разумно рассуждать? (Дж. Маккарти, 1984) и тест на поверхностный «здравый смысл»: может ли компьютер банально рассуждать? (Д. Деннет, 1984). Экспертные системы не способны отвечать ни на глупые вопросы, ни на вопросы, требующие глубокого понимания предмета.

3) Тест «Китайская нация»: возможна ли компьютерная реализация общественного сознания? (Н. Блок, 1978, 1981). Глобальный искусственный интеллект как сингулярная сущность невозможен, он возможен лишь как сочетание деятельности многочисленных разработчиков-людей, каждый из которых обладает уникальным естественным интеллектом.

4) Тест «Китайская комната» (тест Серля): может ли компьютер понимать? (Дж. Серль, 1980 г.). Компьютер понимать не может, более того, мешает процессу «понимания». Он не способен к интенциональности и является всего лишь формальной «тенью» активного осмысления предметной области, на что способен только человек в силу биологического прошлого.

5) Параноидальный тест: может ли компьютер разумно функционировать, как нормальный, психически здоровый человек? (К. Колби, 1971). Нет, не может. Экспертные системы способны имитировать только психические отклонения, паранойю. Параноидальные представления – это идеи программиста, границы которых непреодолимы.

6) Субкогнитивный тест (тест Френча): может ли компьютер обладать «подсознанием»? (Р. Френч, 1990). Компьютер должен «жить» среди людей, тогда его «база знаний» будет наполняться ассоциациями повседневного опыта. При таких условиях он будет проходить тест.

7) Эмоциональный тест: может ли компьютер любить (испытывать эмоции)? (А. Сломан, 1998 г., Дж. Маккарти, 2001). Компьютер

способен «любить», но лишь по-функционалистски, без эссенциальных притязаний.

8) Инвертированный тест: может ли компьютер приписывать ментальность х-системам? (С. Ватт, 1996 г.). Компьютер не может приписывать ментальное, у него отсутствует способность к «воображению». Как только это обнаруживается, значит, мы имеем дело с человеком, но не с компьютером.

9) Гендерный тест: может ли женщина (мужчина) мыслить? (Ю. Генова, 1994). Изучает вопросы гендерной детерминации интеллектуальной деятельности. Включает тест Хопкинса (Р. Хопкинс, 1998 г.): может ли компьютерным образом организованная система испытывать сексуальное влечение и наслаждение? Поднимает вопросы алгоритма работы «секс-машины».

10) Креативный тест (тест Лавлейс): может ли компьютер творить? (С. Брингсфорд, 2001). Компьютер не может творить. Да, созданы искусные программы генерации высокохудожественных литературных, изобразительных и музыкальных произведений. Однако они «творят» в контексте известного.

11) Тест Гёделя-Лукаса-Пенроуза: способен ли компьютер к самоописанию и самоорганизации? (Дж. Лукас, 1961; Р. Пенроуз, 1989). Компьютер не способен ни к самоописанию, ни к самоорганизации. Впрочем, человек тоже на это не способен.

12) Индуктивный тест (тест Мура) как достаточное логическое условие искусственного интеллекта. Это условие индуктивной аргументации в пользу возможности построения искусственного интеллекта. Оно достигается путем коллекционирования программ, успешно прошедших тест Тьюринга, и отбраковки программ, не прошедших этот тест (Дж. Мур, 1976).

13) Новый тест Тьюринга (тест Блока): может ли компьютер осмысленно отвечать на любой синтаксически корректный вопрос? (Н. Блок, 1980). Прохождение теста Тьюринга — необходимое логическое условие искусственного интеллекта. Условие достижимо, но для этого надо включить разработчика в состав компьютерной системы. Но в этом случае мы получим не «искусственный интеллект», но лишь формализованный интеллект разработчиков, т.е. то, что называется «интеллектуальной системой».

14) Кибериадный тест: может ли компьютер творить искусственные миры? (Дж. Баресси, 1987). Предполагается иерархия миров, в которых действуют компьютерные законы создания «себе подобных».

Окрашивает некоторые религиозоведческие вопросы компьютерными коннотатами.

15) Тотальный тест (тест Харнада): может ли компьютер стать неотличимым от человека? (С. Харнад, 2001). Допустим, можно создать компьютерный дубликат человека, то есть реализовать его точную макро- и микрофизическую копию. Однако будет ли дубликат мыслить и осознавать? В свете этого вопроса все современные достижения искусственного интеллекта – лишь игрушечные подделки. Перспективно применение этого теста в конвергентной методологии НБИКС. «Самый тотальный тест» преследует более радикальные креационистские цели и изучает вопросы прохождения компьютером пути социо-когни-био-эволюции и творения компьютером иных компьютерных миров (П. Швайзер, 1998).

16) Нейрокомпьютерный тест или «тест Тьюринга с мозгом» (Черчленды). Классический компьютер не может мыслить, нейрокомпьютер – может. Для этого требуется программа элиминации нейрофизиологических терминов относительно нейрокомпьютерных, как ранее были элиминированы термины народной психологии относительно нейрофизиологических терминов.

17) Тест искусственной личности: может ли компьютерная система стать личностью? Различаются проекты: искусственная личность как антропоморфный робот (Д. Деннет, 1994; Дж. Поллок, 1995) и как экспертная система, репрезентирующая надчеловеческие «знания» (А. Алексеев, В. Деев, 1994).

18) Тест философских зомби: можно ли имитировать сознательное существо посредством бессознательной системы? (Р. Кирк, 1971). Имеет много субтестов. Например, тест Чалмерса: могли ли я быть компьютерным зомби? (Д. Чалмерс, 1995). Для этого нужен компьютер, построенный в соответствии с «новой теорией информации». Что это за теория? Компьютерная реализация философских зомби – это всего лишь компьютерная экспликация художественных образов. Проект искусственной личности, включающий блок «псевдосознания», более правдоподобен (Дж. Маккарти, 1998).

19) Интроспективный тест: может ли компьютер обладать «иным» сознанием? Компьютерная система способна обладать психикой нечеловеческого рода (А. Клифтон, 2003). Более того, глобальный искусственный интеллект (у него-то явно «иное» сознание) обладает настоящим сознанием, а все люди – зомби, так как не владеют полнотой сознательного опыта (Д. Деннет, 2001). Поэтому целостность

когнитивной жизни не в естественном биологическом прошлом человека, как считает Дж. Серль, но в искусственном технологическом будущем человечества, как это доказывает Д. Деннет.

20) Интерактивный тест (В. К. Финн, 2009): может ли компьютер работать с «идеями» как с бесконечно становящимися понятиями, т. е. как с философскими категориями? Тест востребован необходимостью рационализации общественной жизни средствами интеллектуальных технологий и предназначен для изучения вопросов усиления естественного интеллекта. Для реализации теста применяется квазиаксиоматическая компьютерная система поиска закономерностей в фактах, генерации гипотез и верификации/фальсификации фактов. Реализуется ДСМ-машиной, про которую часто говорил Виктор Константинович Финн на заседаниях семинара «Философско-методологические проблемы искусственного интеллекта». Оригинальный тест Тьюринга с его априори разработанной программой не нужен для технологии интеллектуальных систем.

21) Комплексный тест: может ли компьютер ВСЁ? То есть всё то, на что способны частные тесты Тьюринга (А. Алексеев, 2006). Предполагается исследование возможности компьютерной реализации когнитивных феноменов самого широкого спектра. Для реализации теста предлагается машина Корсакова-Тьюринга, которая совмещает коннекционистскую и символьную парадигмы представления «знаний» интеллектуального компьютера.

Все эти тесты по отдельности интересны. Но в сочетании с другими тестами эффект получается, как говорят системотехники, эмерджентным. Например, возьмем тест № 2 (на «здравый смысл»). Интересно отметить, что эти тесты предложены в статьях, которые были опубликованы почти одновременно. Авторы их – крупнейшие когнитивные ученые – Дж. Маккарти и Д. Деннет. Первый вообще является основателем науки «искусственный интеллект». Они критикуют экспертные системы в самый разгар их завораживающих успехов. Об этих успехах вчера говорил Олег Петрович Кузнецов. Однако имеются ли успехи, или это очередной технологический пиар? Предмет тестирования, казалось бы, одинаков – «common sense». Когда Д. Деннет задает экспертной новостной системе SYRUS вопрос, может ли Сайрус Вэнс (тогдашний генеральный секретарь при президенте Картере) завязывать шнурки, то компьютер не отвечает на этот пустяк, хотя досконально «знает» где, с кем, когда и по какому поводу был Сайрус Вэнс. Дж. Маккарти предлагает проверить

медицинскую систему MYSIN с позиции высококлассного врача и обнаруживает, что настоящих-то знаний нет, пациент точно умрет, если доверится компьютеру. «Нам нужны знания гениального врача, а не сведения из медицинских энциклопедий и учебников. А эти знания в компьютер не загнать», – как бы в связи с маккартиевской критикой говорила Т. В. Черниговская год назад на симпозиуме в честь 10-летия НСМИИ РАН [3].

К сожалению, нет возможности подробно останавливаться ни на этих тестах, ни на других тестах, ни на интересных эффектах их совместного применения. Но это уже – комплексный формат.

Идеи изучения частных тестов Тьюринга в комплексном формате

Вопрос построения комплексного теста возник, на мой взгляд, через 50 лет после оригинальной статьи, в 2000 году. Во многом этому способствовал факт того, что срок предсказания Тьюринга по поводу прохождения теста вроде бы наступил, однако никаких праздничных салютов по этому поводу не прозвучало. Тем не менее, с другой стороны, очевидным стал факт того, что была осуществлена компьютерная реализация многих интеллектуальных функций (игры, распознавание/синтез образов и речи, экспертные системы, робототехника и пр.). С некоторой натяжкой можно считать, что предсказание Тьюринга стало сбываться. Возрос и интерес к его тесту.

Пожалуй, впервые идея создания систематической коллекции из частных тестов Тьюринга прозвучала в работе Эйзы Сэйгин, Илиаса Сисекли и Варола Экмана «Тест Тьюринга: 50 лет спустя» [8]. Статья была опубликована в широком доступе, а не в коммерческих скопусах, и поэтому имела широкий общественный резонанс; по крайней мере, в последующих работах, посвященных тесту Тьюринга, на нее, как правило, ссылаются. Авторы представили хороший исторический обзор из десяти тестов, компактно изложили основные вехи в развитии идеи теста Тьюринга. Статья является аннотированным пересказом содержания частных тестов, обрамленным общими вопросами типа: трудно или просто пройти тест Тьюринга, является ли тест операциональным определением естественного интеллекта или не является таковым и пр.

Идеи данной статьи существенно повлияли на проект молодежной конференции «Философия искусственного интеллекта», которая прошла 20 апреля 2004 года в Московском институте электроники и математики (руководители проф. Георгий Петрович Путилов, Андрей Алексеев, Елена Демидова). В ее работе приняли участие

проф. Д. И. Дубровский, проф. К. К. Колин. Было представлено около 80 тезисов докладов, многие из которых опубликованы [10]. Доклады, в основном, были сделаны на основе осмысления работ, на которые навели авторы выше обозначенной статьи.

Цель конференции состояла в том, чтобы погрузить идеи частных тестов в философский, научный, исторический и технологический контекст. Конечно, такой контекст должен быть максимально близок проблемам искусственного интеллекта. Поэтому предметом изучения явились вопросы обусловленности функционалистской концепции мышления тестом Тьюринга, связь идей Алана Тьюринга как основателя философии искусственного интеллекта с идеями Джона Маккарти как основателя науки «искусственный интеллект», практические приложения теста в контексте технологии интеллектуальных систем. Так, на секции «Функционалистская концепция мышления. Тест Тьюринга: pro et contra» звучали темы: «Базовые положения теста Тьюринга», «Может ли машина мыслить? Полемический стандарт Тьюринга», «"Искусственный интеллект" Алана Тьюринга», «Многообразие тестов Тьюринга», «Тест Френча: субкогнитивистское опровержение концепции Тьюринга», «Тест Блока: антибихевиористское опровержение тьюринговой концепции мышления», «Тест Блока: нестандартные антибихевиористские возражения Тьюрингу», «Тест Лавлейс: машина творить не может!», «Наивная психология и инвертированный тест Тьюринга», «Тест Серля: интенционалистское опровержение концепции Тьюринга», «Критика сильного искусственного интеллекта. Аргумент Гёделя», «Тест Тьюринга и Д. Деннет», «Компьютер может мыслить (М. Минский)», «Социокультурные аспекты теста Тьюринга», «Тест Тьюринга и паранойя». Вопросы практического программирования игры в имитацию изучались в докладах «Зачем вкладывать деньги в тест Тьюринга? Лойбнеровская премия», «Программирование Теста Тьюринга». Ряд докладов акцентировал внимание на тесной связи идеи теста Тьюринга с машинным функционализмом Х. Патнэма: «Функционализм, физикализм и искусственный интеллект», «Парадигма функционализма: как представить ментальное в нементальных терминах?», «Тьюринг и проблема вычислимости сознания», «Квалиа и парадигма функционализма», «Функционалистский статус любви».

Теоретические приложения теста звучали в преломлении с идеями Дж. Маккарти: «Что такое искусственный интеллект?»; «Зачем искусственному интеллекту философия?», «К вопросу об

эпистемологической адекватности репрезентаций», «Псевдоволья»; «Логические аспекты создания искусственного интеллекта», «Реализационные перспективы теории речевых актов», «Экспертные системы, основанные на здравом смысле», «Проблема дискурса искусственного интеллекта. Конструкторская позиция», «Конкретизация термина "искусственный интеллект"».

Историко-философская проблематика в связи с тестом Тьюринга звучала в докладах «Контентная модель смысла (А. Ф. Лосев)», «Виртуальная реальность и математика Н. Кузанского», «Искусственный интеллект и стоическая эпистемология», «Проблема идеального и искусственный интеллект», «Проблема естественных видов в искусственном интеллекте», «Интуитивистские ориентиры моделирования смысла (А. Бергсон)» и даже прозвучал доклад «Искусственный интеллект и святоотеческий опыт».

К практическим приложениям теста Тьюринга были отнесены не попытки создания программ, играющих в игру в имитацию, но более практические цели: «Бионика как направление робототехники», «Квантовые компьютеры и квантовая механика», «Парадигма коннекционизма как методология нейрокомпьютерной технологии», «Методологические аспекты нанотехнологии», «Естественно-языковой интерфейс: три подхода к моделированию "смысла"», «Человек и компьютер: друзья или враги?», «О распределении функций между человеком и компьютером в информационно-коммуникационных технологиях», «Клоны и «полуискусственный интеллект»; «"Квалиа" как базовая категория виртуалистики», «Что значит "Быть роботом"»? Доклад «Зомби и искусственный интеллект» в контексте изучения когнитивных характеристик колоний антропоморфных роботов совместил два вопроса 1960-х годов: «Может ли машина мыслить?» и «Есть ли жизнь на Марсе?».

Обзор этой конференции 2004 года в сегодняшнем моем докладе специально сделан для того, чтобы показать, сколь разнообразные и интересные идеи непосредственно и опосредованно порождаются проблемами комплексного осмысления идей теста Тьюринга. Этот тест задает принципиальные и простые положения, которые без особых концептуальных проблем инкорпорируются в философско-методологический, социокультурный, исторический и технологический контексты. Массовый интерес вызывает тематика теоретических принципов искусственного интеллекта, но отнюдь не программно-инженерного «мелкотемья» проектов интеллектуальных программ.

В конференции участвовало свыше 250 человек, в основном, студентов старших курсов МИЭМ. Ребята искренне работали над философскими текстами. Попробуйте заставить современного российского студента поработать ради учебы, ведь он постоянно где-то работает, чтобы иметь возможность оплатить эту учебу! А тема теста Тьюринга «запала в душу» студентам – прикладным математикам и программистам.

Вернусь к англоязычным исследованиям. В 2008 году вышла в свет самая фундаментальная на сегодняшний день книга по тематике тьюрингового тестирования: «Анализ теста Тьюринга: философские и методологические исследования проблематики мыслящих компьютеров» [9]. Ее объем свыше 500 страниц. Книга была издана в скопусовском Шпрингере и поэтому менее известна, нежели отмеченная выше статья Э. Сэйгина и др.: на книгу не так много ссылок. Редакторами книги явились организаторы лойбнеровских игр: вице-президент этих состязаний социолог Роберт Эпштейн (он имеет степень PhD), программист Гарри Робертс и менеджер, учительница английского языка Грэйс Бебер. Несмотря на невысокий ученый статус редакторов, в книгу вошли статьи крупнейших когнитивных и компьютерных ученых: Эндрю Ходжеса, Неда Блока, Джона Лукаса, Селмера Брингсйорда, Пола Черчленда, Джека Копленда, Джона Серля, Жана Лассажа, Хьюго Лойбнера, Марка Хамфриса, Дугласа Лената, Стюарта Ватта, Рэя Курцвейла и др. Из бывших «наших» представлена статья Евгения Демченко (украинца, живущего в России) и Владимира Веселова (русского, живущего в Америке). Даже Ноам Хомский, пребывающий в весьма почтенном возрасте, посвятил пару страниц тесту Тьюринга. Была переиздана оригинальная статья Алана М. Тьюринга [1] с объемными постраничными сносками, детально описывающими мнения того или иного исследователя по различным пунктам игры в имитацию.

На мой взгляд, это очень крупная антология. Тем не менее, в книге отсутствует идея интеграции этих тестов. Пожалуй, только в работе Роберта Хорна «Тест Тьюринга: сопоставление и обсуждение дискуссий» была представлена идея интеграции тестов в виде диаграмм идей по поводу теста и идей по поводу этих идей. Материалы демонстрировались в виде фотографий, картинок, рисунков, стрелок и пр. По всей видимости, художественные образы облегчают восприятие материала. Однако это затрудняет реальную аналитическую работу с частями частных тестов и частями этих частей. То есть автор

создал некоторый путеводитель по тестам (нечто подобное этому иногда почему-то называют «дорожной картой проекта»). Однако он не уделил внимания вопросу – как и зачем эти тесты можно использовать совместно. И в целом, в этой громадной коллективной монографии не продуманы основания для классификации частных тестов. Поэтому невозможна их последующая систематизация, унификация, интеграция.

Аналогичная по замыслу работа – комплексно осветить историю идеи теста Тьюринга – была проделана чуть позже организаторами научно-практической конференции «Тест Тьюринга: философские интерпретации и практические реализации», которая состоялась на философском факультете МГУ имени М. В. Ломоносова в октябре 2010 г. Ее организаторами и редакторами последовавшего сборника [3] явились заведующая кафедрой философии языка и коммуникаций философского факультета МГУ, к.ф.н. Анна Анатольевна Костинова и к.ф.н. Наталья Юрьевна Ключева. Конференция была посвящена шестидесятилетней годовщине публикации статьи А. Тьюринга. С докладами выступили проф. В. В. Васильев («Показывает ли "Китайская комната" Серля неэффективность теста Тьюринга?»), проф. В. К. Финн («Отношение к Тесту Тьюринга с точки зрения современных интеллектуальных систем»), к.ф.н. А. Ю. Алексеев («Комплексный Тест Тьюринга»), к.ф.н. Н. Ю. Ключева («Методологическая ценность Теста Тьюринга для развития компьютерных наук»), А. Потапов («Искусственный интеллект: представимость и возможность»), к.ф.н. Т. А. Вархотов и др.

И еще одно из последних мероприятий раскрывает интересные аспекты комплексного изучения теста Тьюринга: 28 мая 2014 г. в Институте философии РАН состоялся круглый стол на тему «Может ли компьютер «читать»?». Мероприятие было посвящено 450-летию книгопечатания на Руси (напомню, что 1 марта 1564 года в Москве вышла в свет первая датированная русская книга «Апостол», изданная Иваном Федоровым и Петром Мстиславцем). На круглом столе обсуждались следующие темы. Комплексный тест Тьюринга на искусство чтения: может ли компьютер «мыслить», «понимать», «творить», «переживать», «осознавать», «воображать» и пр.? Компьютерная герменевтика: перспективы перехода от систем моделирования «данных» и «знаний» к системам моделирования «смыслов». Электронная книга в переходный период от традиционной к электронной культуре. Антропологические проблемы электронной культуры: может ли

читать «человек-компьютер»? Перспективы разработки нейроинтерфейсов электронной библиотеки. Проблема «понимания» текстов в интеллектуальной робототехнике. Может ли «понимать» компьютерная система машинного перевода?» и др. Несомненно, это – актуальная тематика для практических исследований искусственного интеллекта и, подчеркну, она возникает в тесной связи с интегральным осмыслением тьюринговых идей.

Наш краткий обзор истории развития идеи теста Тьюринга как комплексной идеи, пожалуй, достаточно убедительно показывает, что с начала 21 века происходит систематическое и всестороннее осмысление проблематики компьютерного «мышления», «сознания», «творчества» и т. д. Сегодня вряд ли можно говорить о создании некоего особого специфического частного теста. Так или иначе, затрагиваются нюансы других подобных тестов. Те тесты, в свою очередь, затрагивают другие тесты и т. д. Получается сложная разветвленная сеть идей, в которой, помимо достаточно четкой сети понятий в схеме определенного теста, наблюдаются гипер-, мета-, инфраструктурные составляющие. Гиперсетевые параметры задают систему ссылок: какой путь приводит нас к той или иной идее конкретного теста, какие авторы, направления, какие ключевые идеи пролагают этот путь. Метатестовые характеристики определяют язык комплексирования разнородных тестов путем типизации и унификации особенностей частных тестов. Один из вариантов такого языка предлагался в [5]. Инфраструктурные составляющие – это те, которые составляют собственный базис тьюринговой концепции. К примеру, если меняется концепция компьютера и мы переходим от проекта вычислимости дискретных элементов к проекту континуальной вычислимости, то понятно, к каким изменениям «цифрового общества» это приводит. Или, скажем, меняется отношение судьбы к компьютерной программе: теперь она должна не имитировать интеллект человека, но, напротив, усиливать его. Тогда и судьба приобретает новые способности, и теперь ему не противостоят терминаторы с симулякрами.

Так как просматриваются гипер-, мета-, инфра-составляющие комплексного теста Тьюринга, то этот тест уже сегодня можно представлять в виде сложной концептуальной системы как необходимой подсистемы проекта искусственного интеллекта. Точнее, комплексный тест Тьюринга должен верховодить другими подсистемами этого проекта.

Что тестирует комплексный тест Тьюринга?

Напомню, что в своей статье А. Тьюринг заменил бессмысленный вопрос «Может ли машина мыслить?» на вопрос «Способен ли компьютер наряду с людьми играть в игру в имитацию человеческого (интеллектуального) поведения?». Однако ограниченность когнитивных проблем искусственного интеллекта сферой мыслительной деятельности представляется слишком узкой. Конечно, в *Homo sapiens* «разум» играет определяющую роль, но этим не исчерпывается антропологическая метрика. Помимо «интеллекта» имеется многое другое.

Поэтому тьюрингов вопрос сегодня мог бы звучать так: «Способен ли компьютер имитировать когнитивное поведение»? Когнитивное поведение понимается достаточно широко. Это – витальные феномены биологических систем: как описать феномен «жизни», отталкиваясь от естественнонаучных законов, закономерностей, регулярностей, алгоритмов. Этими исследованиями занимаются специалисты проекта «Искусственная жизнь». Во-вторых, когнитивные феномены изучаются в психофизиологическом формате. Это собственное значение слова «когнитивный». Такие исследования проводятся в нашумевших национальных и межнациональных проектах «Искусственного мозга». Далее, когнитивные феномены рассматриваются в контексте проекта «Искусственная личность»: как, исходя из выявленных ментальных и, шире, психологических регулярностей и репрезентаций, компьютерно представить самость, Я, свободу или, скажем, феномен личного морального вменения? И, наконец, когнитивные феномены, которые изучаются в проекте «Искусственное общество», связаны с вычислительными аспектами «социального» как надличностными параметрами общности, дружбы, вражды, справедливости, социального идеала и пр.

На мой взгляд, очень удачный термин для этих проектов «об искусственном» в 2005 году предложил первый председатель нашего НСМИИ РАН В. Л. Макаров. Он сформулировал интегральный проект «Искусственный мир» и обозначил ряд эпистемологических проблем построения компьютерных имитационных моделей этих самых различных когнитивных феноменов [10].

Комплексный тест Тьюринга, собственно, и тестирует когнитивные феномены этого мира «искусственных миров».

Является ли статья Тьюринга философским исследованием?

Так как сегодня у нас конференция посвящена философии искусственного интеллекта, и изучаемая статья – начало этих исследований, то правомочен вопрос – а сама статья из какой сферы? Вроде бы никаких онтологий с гносеологиями в ней нет. Однако статья А. Тьюринга уже сейчас оказала колоссальное значение на современную науку, культуру, технологию. Это влияние произошло, конечно, за счет ее философского содержания. В статье четко прозвучала идея вовлечения людей разных специальностей в междисциплинарные исследования по построению «мыслящих машин». Это требует методологической координации и интеграции труда программистов, математиков, филологов, психологов, биологов и др. Искусственный интеллект – это, по определению, сфера междисциплинарных исследований. Встает вопрос – кто в этой системе взаимодействия специалистов главный? Конечно, философ, – по крайней мере, согласно классическому определению специфики этой меж- (мета-, транс-) дисциплинарной профессии. Однако кто это такой – современный когнитивный философ, осмысливающий первопринципы компьютерной реальности? Ведь это не только изучение фундаментальных отношений «Я-мир» (точнее, «Я-искусственный мир»), но и исследование когнитивных предпосылок и способностей как условий исследования этих когнитивных предпосылок и способностей.

Небезынтересно отметить, что А. Тьюринг философом себя не считал и часто подчеркивал, что он не кто иной как математик. На это заявление указывал Эндрю Ходжес, известный биограф А. Тьюринга. Однако тот же Э. Ходжес в 1997 году издает книгу «Алан Тьюринг» в известной серии «Великие философы XX века». Кстати, Виктор Антонович Садовничий в начале своего доклада сказал, что – не философ, но математик, тем не менее, с философией его сближают вычислительные проблемы искусственного интеллекта. Предложенная им тематическая рубрика «Суперкомпьютер и искусственный интеллект» чрезвычайно значима для развития философии искусственного интеллекта, так как, например, она связана с проектом сверхчеловеческой вычислимости (ведь человек – это последовательно мыслящее существо, а суперкомпьютер работает параллельно). В этой рубрике вряд ли целесообразно использовать какие-то метафизические категории. Несмотря на их отсутствие, ее тематика имеет философское, методологическое значение: она междисциплинарна.

В связи с этим интересно высказывание автора субкогнитивного теста Тьюринга, психолога Роберта Френча, о том, что Алану Тьюрингу пришла философская известность за то, что, исследуя «разум», он не попал в болото философских спекуляций по проблемам «разум/тело». То есть всего того, чем сегодня занимаются международные и отечественные «Центры исследования сознания». Может быть, сила когнитивного философа состоит в преодолении искушения словесных хитросплетений метафизической когнитивистики по поводу возможности/необходимости/истинности/свободы и пр.? По проблеме сознания написаны сотни тысяч работ. Не лучше ли все это жестко стереть с носителей информации, как это раньше делал Юм, сжигая философские трактаты в камине?

Помимо антиметафизических интенций статьи первого философа искусственного интеллекта, важно подчеркнуть, что в этой работе были затронуты глубинные мировоззренческие вопросы, присущие каждому человеку: о разуме, сознании, свободе, личности. Эти вопросы возникли в контексте концепции цифровой механической вычислимости. Напомню, что А. Тьюринг признан одним из авторов этой концепции: его формальная теория алгоритма 1936 года емко и обобщенно описывает жизнь компьютерного мира, в котором мы все сейчас живем. Конечно, практически эту теорию не применить – ведь никакого смысла нет в том, чтобы диалог с компьютерным собеседником программировать инструкциями машины Тьюринга. Это теоретически и принципиально можно, но практически не нужно. Главное для философской рефлексии – выявить обобщенный теоретический принцип.

Однако главная заслуга статьи Тьюринга, на мой взгляд, все-таки в другом. За всю историю человечества трансцендентное искушение «творения себе подобного» не получало столь конкретных реализационных оснований, как это было предложено в данной статье. Глина людей ветхозаветного бога, мрамор статуи Пигмалиона, химико-электрические процессы демона Франкенштейна – все это сказочные фантазии на тему искусственной личности. У А. Тьюринга – не сказки. Очень убедительно показаны перспективы применения программируемых компьютеров. И даже дата появления компьютерных интеллектуальных двойников совпадает с точностью до десятка лет: 2000-ый год. Ведь если совсем недавно компьютерный проект Тьюринга вызывал споры по поводу полноты компьютерной аппроксимации когнитивных компетенций человека, то сегодня вряд ли кого-то надо

убеждать в этих перспективах. «Будущее уже наступило» – написано на лозунгах некоторых компьютерных фирм. Тьюринговые игроки обыгрывают людей в интеллектуальные игры, создают шедевры, разговаривают на разных языках, принимают рациональные решения, навязывают собственные убеждения и, имитируя человека, напрочь вытесняют его из традиционных сфер труда.

Идеи статьи А.Тьюринга породили специфическое и, не побоюсь сказать, главное направление философских исследований искусственного интеллекта. Условно его так и можно обозначить – «тьюрингово тестирование». В рамках этого направления обсуждаются принципиальные вопросы: что могут и не могут компьютеры; какие когнитивные феномены компьютеры имитируют, репрезентируют, репродуцируют, рекреацируют; как и посредством чего компьютеры так делают либо почему пока этого не делают либо принципиально не могут этого сделать. И, конечно, тест Тьюринга позволяет рационально обсуждать истинные практические вопросы, связанные не с оптимизацией программных систем, но с социокультурными последствиями их применения: зачем все это надо?

Согласитесь, что любой исследователь искусственного интеллекта перед началом своей работы должен быть убежден в том, что *компьютер не может, может либо сможет мыслить*. А эти модусы убежденности формируются в контексте тьюрингова тестирования. Поэтому-то *тест Тьюринга представляется необходимым условием исследований искусственного интеллекта*, причем необходимым условием в строгом логическом смысле: если отсутствует явная или неявная ссылка на тест Тьюринга, то это исследование уже не относится к искусственному интеллекту. Например, некоторое исследование представляет программное воплощение важных нейрофизиологических опытов вместе с конкретным приложением психологической теории. Вроде бы такой продукт подпадает под тему «искусственный интеллект». Так программисты и поступают: излагают некоторую теорию на языке программирования. Однако если при этом нет указания, какие параметры коррелируют с приватными специфическими когнитивными феноменами пользователя этой программой, если не проводится междисциплинарное обсуждение возможностей и социокультурных последствий такой реализации, то тогда это исследование – из какой-то другой сферы знаний и практик, но не из области искусственного интеллекта.

Роль комплексного теста Тьюринга не ограничивается искусственным интеллектом. Всесторонне и «умно составленный» (comprehensive) комплексный тест Тьюринга способен играть роль *концептуального инструментария познания и понимания когнитивных феноменов компьютерного мира, мира электронной культуры*. Никакой частный тест, будь он самый совершенный, не способен к этому, слишком уж сложна феноменология этого компьютерно фундированного мира. Поэтому философская значимость теста Тьюринга, трактуемого исключительно в комплексном формате, проявляется в способности целостно охватить и когнитивные и компьютерные явления разом и изучить эти явления в эпистемологическом, онтологическом и праксеологическом контекстах.

Эпистемологические приложения комплексного теста

В эпистемологическом плане, выполняя интеррогативную функцию познания, комплексный тест Тьюринга позволяет поднимать вопросы: «Может ли компьютер мыслить, понимать, жить, творить, любить, осознавать, быть личностью, обществом, обладать свободой и пр.?» Короче, может ли компьютер делать как бы «всё»? Получая ответы на данные вопросы, комплексный тест выполняет функцию дефиниции «мышления», «понимания», «творчества», «сознания», «другого», «личности», «Я» и пр. При этом исключается концептуальный дискомфорт употребления компьютерно-ориентированных понятий: «машинное мышление», «компьютерное сознание», «человек-компьютер», «любовь робота». Ведь в комплексном тесте, как и в оригинальном тесте, определяются *x*-феномены *y*-систем, *z*-субстраты которых не существенны или вовсе не известны.

Еще одной важной функцией рационального познания компьютерного мира является критика проектов электронной культуры. Например, легко доказывается то, что современные компьютерные интеллектуальные технологии не относятся непосредственно к сфере исследований искусственного интеллекта. Хотя многие известные фирмы, специализирующиеся в этой области, серьезно рекламируют свою продукцию как «технологии искусственного интеллекта». Например, программы известной фирмы АBBYU «фактически понимают тексты», – так полагает руководство этой фирмы. То есть компьютеры, в самом деле, может быть и не так, как человек, но фактически способны понимать. Странно то, что эксперты таких фирм не интересуются фундаментальными вопросами искусственного интеллекта: Тесту Серля в этом году исполнилось, например, 36 лет

и именно в концепции «Китайской комнаты» четко прозвучала критика таких «понимающих» программ.

Схожим образом, путем использования ряда частных тестов, критически оцениваются современные социальные проекты, в которые вовлекается идея компьютеризации когнитивных феноменов. Предельно очевидно комплексный тест Тьюринга оценивает «глупость» государственных решений калькуляционной оценки когнитивной сферы: подготовленность абитуриента оценивается автоматизированным ЕГЭ, знания студента – средствами федерального интернет-экзамена, авторитет преподавателя и ученого – индексами цитирования, которые вычисляются «сырыми» программами электронных библиотек. Компьютеризация затрудняет познание, отдаляет понимание – как предмета понимания, так и понимания того, понимает ли х-система этот предмет. Перспективы цифрового общества, в котором калькуляция когнитивной сферы возведена в социальную норму, ярко раскрываются тьюринговым тестом зомби.

В-третьих, комплексный тест Тьюринга исполняет функцию конструирования новых принципов компьютерной технологии. Например, позволяет обосновать стратегию переориентации технологии искусственного интеллекта от проблематики моделирования иллюзорного феномена «сознания» на решение прагматичных задач моделирования феномена «творчества». Такой ориентир мне представляется направлением «главного удара» в текущей доктрине развития искусственного интеллекта. Ну не способны современные концепции вычислений расшифровать код Дубровского.

В прикладном плане комплексный тест способствует разработке новых принципиальных проектов компьютерных систем. Категориальный статус таких проектов сравним с машиной Тьюринга. Например, машина Корсакова-Тьюринга на теоретико-алгоритмическом уровне соединяет коннекционистскую и символическую парадигмы искусственного интеллекта. Это позволяет кардинально реформировать современный компьютеризм и в конечном итоге изменить облик электронной культуры. В таких алгоритмах равноправные роли получают «математик» (метафора машины Тьюринга) и «художник» (метафора машины Корсакова).

Комплексный тест – это концептуальный инструментальный «чистых» дискуссий в познании феноменов электронной культуры. К аргументации не примешиваются технологические, идеологические, политические, экономические, религиозные, морально-этические,

эстетические и иные соображения. Такая форма познания компьютерного мира в финансовом плане стоит ровно столько, сколько стоит «мышление», а сегодня в нашей стране подобный абстрактный труд стоит меньше труда уборщицы или, скажем, охранника. Однако в когнитивистском отношении комплексный тест Тьюринга провоцирует крупные мысленные сражения, которые позволяют сберечь многие миллионы долларов при реализации проектов «искусственных мозгов». Эти войны разгораются в сфере когнитивной философии, точнее, в методологии компьютерного функционализма.

Машина Тьюринга и современный функционализм

Современный функционализм как философская, естественнонаучная, социогуманитарная и технологическая парадигма имманентно инкорпорирован в исследования психофизиологической и психотехнологической проблематики. Поэтому его исконные источники не замечаются. В самом деле, они очень далеки от философско-эпистемологических проблем изучения познавательной и психической деятельности, так как лежат в сфере математического обоснования теоретико-алгоритмических проблем и инженерных проектов построения электронно-вычислительных машин.

В работе 1936 г. А. Тьюринг предлагает инженерную формализацию интуитивного понятия «алгоритм», которая с 1937 года называется «машиной Тьюринга». Машина состоит из управляющего устройства, которое посредством считывающей/записывающей головки манипулирует символами на ленте и переходит по множеству дискретных состояний в соответствии с автоматной таблицей, репрезентирующей программу, реализующей алгоритм. Способ функционирования машины – это метафора работы математика. Чтобы сравнить преимущества машинного, инженерного подхода к формализации понятия «алгоритм», выделим математический (А. Чёрч) и лингвистический (А. А. Марков – мл.) подходы. Трактовки алгоритма эквивалентны, что утверждается без доказательства в знаменитых тезисах Тьюринга – Черча, Маркова – Тьюринга и, в силу эквивалентности, в комплексном тезисе Тьюринга – Черча – Маркова. Разные трактовки утверждают то же самое. Однако машина Тьюринга – на вершине эвристической демонстративности, так как, во-первых, легко вообразима в формате мысленного примера – её можно концептуально «пощупать» и «поработать» с ней, в отличие от других подходов, которые требуют усилий для представления средств реализации математических функций либо операций с текстом. Демонстративность

машины Тьюринга, по всей видимости, главное достоинство и среди других подходов к определению понятия «алгоритм». Например, в колмогоровском подходе алгоритм – это конструктивная работа по конструированию конструктивных математических объектов. Однако непонятно, какое именно устройство способно таким образом работать. Кстати, А. Тьюринг в работе [1] задавался не только вопросом, что такое «интеллект», но и что такое «машина». В самом деле, что это такое? А что такое «компьютер»?

Во-вторых (и это – главное), машина Тьюринга двуаспектно фиксирует связь между формальным (символы на ленте) и каузальным (устройство управления). Возможность такого отношения способствовала идее Х. Патнэма отождествить психические состояния человека с логическими состояниями машины Тьюринга, а физиологические состояния с физическими состояниями технического устройства. Анализ психических состояний ведётся от первого лица либо от третьего лица при условии эмпатического сопереживания, т.е. изучаются приватные когнитивные феномены. Динамика психических состояний и смена логических состояний машины Тьюринга аналогичны. Функционирование мозга с одной стороны и «компьютера» с другой, так же аналогичны. Логические состояния реализуются физическими процессами электронных схем, тем самым постулируется каузальная связь психического (логического) и физического (технического).

Из этого следуют три классических функционалистских тезиса: 1) Эгологический тезис: Я (другой) – машина Тьюринга и мои (его) психические состояния тождественны логическим состояниям машины Тьюринга (1960 г.) [12]; 2) Тезис множественной реализации: психические состояния могут быть реализованы человеческим мозгом, компьютером, тинной марсианина, если они имеют функциональную организацию, подобную функциональной организации психики сознательных существ (1967 г.) [13]; 3) Тезис ментальной причинности: каузальные отношения между когнитивными феноменами – это функциональные отношения.

Тезисы достаточно сильные. С 1980-х гг., пытаюсь совместить машинный функционализм с социокультурной проблематикой и с достижениями естественных наук о мозге, Х. Патнэм снимает эти заявления. Солипсическое решение вопроса множественной реализации возникло из-за того, что ментальные состояния – это мои личные состояния, укоренённые в социальных отношениях между людьми. Приватные когнитивные феномены невозможно

ни репрезентировать машинными состояниями, ни описать квази-алгоритмическими дескрипциями и прескрипциями. Сложны и необъяснимы каузальные связи между социальным, психическим и биологическим.

В параграфе «Компьютерная психология и теория интерпретации» в [14] (1983 г.) Хилари Патнэм принципиально утверждает, что когнитивная теория не способна стать теорией ментального содержания, в основном, из-за проблем реализации. Физическая реализация психического, даже при наличии опосредующих функциональных систем между ними, т.е. построение своего рода «психологической физики», представляется утопией, наподобие того, как и социальная физика О. Конта оказалась лишь мечтой, – пишет он в 1997 году в работе «Функционализм: когнитивная наука или научная фантастика?» [15, Р.47]. Собственно, и когнитивная наука как таковая несостоятельна, так как более-менее прочный базис создаётся функционализмом. Другие психофизические парадигмы даже и не достигают методологических границ, гарантированных функционализмом. Поэтому наука бессильна в разгадке когнитивных тайн человека: как возможны психика, сознание, Я?

В связи с этим было бы интересно применить критические положения Х. Патнэма к теории когнитивных вычислений, прозвучавших в докладе К. В. Анохина. Как-то очень напоминает «когнитивный компьютеринг» патнэмовскую версию «психологической физики»: гиперсети из когов, локов – это как шарики на ниточках. К сожалению, нет возможности подробно на этом останавливаться, отмечу только, что подобные вопросы провоцируют мысленные баталии в сфере методологии компьютерного функционализма и новейшая психофизическая теория – теория когнитума – заслуживает мощных концептуальных сражений.

Сегодня, в условиях развития НБИКС, антисциентистские настроения Х. Патнэма непопулярны. Активно эксплуатируются ранние оптимистичные идеи машинного функционализма. Компьютеризм (computationalism) обогатил оригинальный функционализм репрезентационизмом и коннекционизмом. Первая парадигма ввела в когнитивную науку контент машинных состояний, интенцию, семантику. Вторая – квазиалгоритмическую парадигму «вычислений», так как чистый нейрокомпьютеринг с трудом вписывается в традиционный формат вычислительных теорий. Компьютеризм пошёл далее решения психофизической проблемы. Проблема стала частным моментом

когнитивистской экспансии на физические, биологические, персональные, социальные объекты: когнитивно-компьютерные системы формально инвариантны относительно субстратных составляющих технологий из комплекса НБИКС, что, собственно, даёт право искусственному интеллекту стать во главе конвергентной методологии данного комплекса. Функционалистские проблемы уникальности множественной реализации и ментальной причинности погрузились в тень радужных проектов типа «Россия-2045». Таким образом, патнэмовский эгологический тезис «Я – машина Тьюринга» не только не утратил силу, но и существенно усилился.

Тестовый компьютеризм

Парадигма функционализма в современной когнитивной философии возникла в 1960-е гг. в ходе критического совмещения достижений бихевиоризма и физикализма. Функционализм подготовил плодородную почву для развития методологии описания, объяснения, предсказания, воспроизведения когнитивных феноменов. Источником функционализма явились идеи искусственного интеллекта А. Тьюринга. По сей день исследования в этой сфере являются самой продуктивной составляющей современной философии, так как снабжают её эмпирическим базисом – компьютерными реализациями когнитивных механизмов, использование которых позволяет изучать старые философские проблемы в новом социокультурном контексте компьютеризованного мира. Об этом говорил В. А. Лекторский еще на первой конференции «Философия искусственного интеллекта» в 2005 году [Лекторский, 2006].

Другие крупные парадигмы когнитивной философии (физикализм, бихевиоризм, спиритуализм, эпифеноменализм, дуализм, дуаспектизм и пр.) вряд ли следует считать надежными эпистемологическими основами. Они способны нарисовать красивые блок-схемы психической деятельности (интеракционизм), предложить новые заманчивые теории информации (неодуализм), возбудить художественные интуиции (эмерджентизм в формате синергетического аутопоэзиса) и пр. Однако в условиях прохождения достаточно сурового фильтра компьютерной методологии и инженерии нефункционалистские парадигмы раскрывают свою несостоятельность. Каузальные зависимости блок-схем лишаются номологического статуса. Натуралистический дуализм ведёт к эпифеноменалистическому отрицанию сознания даже у человека, не говоря уже о животных и искусственных системах. Синергетика, которая якобы

сменила кибернетику в плане объяснения и конструирования социально-технической реальности, пользуется настолько невнятно артикулируемым словарём, что попытки его лексического разбора дают повод усомниться в том, что синергетик сам понимает собственные утверждения.

В отношении главных оппонентов – физикализма и бихевиоризма – функционализм предлагает комплексный формат изучения психофизической реальности. Умеренный функционализм утверждает, что когнитивный феномен психического содержания (боль, мысль, желание, убеждение) зависит не только от внутреннего состава и строения системы (физикализм) и не только от внешней поведенческой диспозиции (бихевиоризм), но и от функции (роли), которую данный феномен выполняет в когнитивной системе в целом. Радикальный функционализм пользуется форматом исключющей дизъюнкции в этой формуле. И самое интересное в том, что оригинальный функционализм, функционализм машины Тьюринга, универсален относительно всех теорий сознания, ведь эти теории прекрасно представимы на лентах машины Тьюринга. И даже сама функционалистская теория там же представима (см. в [13]).

Современный функционализм как парадигма когнитивной философии многообразен и вариативен. Невозможно говорить о некотором едином «функционализме» с чётко заданным словарем, методологией, программой. Следует говорить лишь о «функционализмах». Причем количество разновидностей «функционализма» равно количеству функционалистов, умноженному на количество интервалов времени, в течение которых у функционалиста сохраняется некоторая устойчивая концепция функционализма.

Показательным примером «относительности теории функционализма от времени» являются различные редакции статьи «Functionalism» Джанет Лёвин (2004 и 2013 гг.), изданных в престижной Стэнфордской философской энциклопедии. В первой редакции (2004 г.) автор выделила следующие разновидности: функционализм машинных состояний, психофункционализм, аналитический функционализм, функционализм тождества функциональных состояний. В редакции 2013 г. [17], Дж. Лёвин немного изменила свои функционалистские взгляды. Первые три разновидности функционализмов остались без особых изменений. Однако вместо четвёртого предложены новые формы: функционализм ролей (Role-functionalism) и функционализм реализаторов (Realizer-functionalism). В принципе,

это объяснимо общемировыми натуралистическими тенденциями в изучении когнитивной проблематики. Однако точно то, что в классификации функционализмов упустила Дж. Лёвин: и в старой и в новой редакции статьи «Функционализм» она, по непонятным причинам, не упоминает компьютерный функционализм (computational functionalism). А компьютеризм, на мой взгляд, это высшая стадия функционализма.

Компьютеризм (computationalism) – самая современная парадигма современной когнитивно-компьютерной науки. Она задаёт парадигму изучения когнитивных феноменов крайне широкого спектра – витальных, ментальных, персональных, социальных. Для компьютеризма встречаются иные транскрипции: «компьюшизм» (Юлина, 2004), «компьютационализм» (Алексеев, 2004), вычислительная теория сознания («The Computational Theory of Mind» (Хорст, 2011 [18]). Все эти «кальки», на мой взгляд, неудобно произносить.

Компьютеризм – это функционализм всех перечисленных разновидностей, каждая из которых конкретизируется тогда, когда супервенторы и/или реализаторы представлены в компьютеризуемом формате. Например, боль как когнитивный супервентор – это не просто статическое психическое состояние или содержание. Это – компонента квазиалгоритмически функционирующей динамической системы, факторы которой задаются ментальными причинами и ролями «боли» в составе функционирования целостного организма. Реализатор «боли» – это компьютер, который трактуется двояко – в узком и широком смысле слова «компьютер». В узком смысле «компьютер» рассматривается в инженерном формате, т.е. со всеми прилагаемыми видами обеспечения вычислительного процесса: техническим, программным, информационным, лингвистическим и пр. Когда «компьютер» трактуется в широком смысле, то, к примеру, мозг, тело, общество рассматриваются как абстрактные машины исполнения алгоритма. Имеется ещё одна трактовка термина «компьютер», достаточно часто фигурирующая в концептуальных построениях современных функционалистов: «компьютер» – это некоторый проект, предполагающий реализацию в будущем, но отнюдь не проект компьютерной машины Бэббиджа – Тьюринга – Неймана; искусственный интеллект на базе старой, дряхлой вычислительной машины – это фантастика дурного вкуса (Д. Деннет). Компьютеризм строго продолжает идею машинного функционализма и является наиболее аутентичной разновидностью функционализма. В нем четко

прописываются: экстерналистская/интерналистская программы изучения супервенции/реализации, проекты «воплощения» когний в компьютерной системе, эпистемологический базис в форме «знаний» когнитивно-компьютерной системы.

В компьютеризме идея функционализма машинных состояний обогащается идеей содержания этих состояний в виде теорий репрезентации и «внутреннего языка мысли» (это положение чётко прописал С. Хорст [18]). Дело в том, что машинный функционализм состояний Х. Патнэма был расширен его учеником Дж. Фодором сразу же после обнаружения проблем с тем, что машинные состояния не позволяют манипулировать с содержанием этих состояний. Скажем, в классическом патнэмовском примере «триггер № 36 включен» отнюдь не означает того, что Я обладаю квалиа, если «Я – это машина Тьюринга». Требуется репрезентация содержания «боли», например, в психофункционалистском формате. Другой пример репрезентационного подхода: если изучаются когнитивные феномены социокультурного содержания, то в «культуре» модульно выделяются блоки «данных», «знаний» и пр. У функционалиста – специалиста в области теории культуры – такая поспешная технологическая структуризация способна вызвать недоумение. Однако для компьютеристского представления «культуры» концептуальных затруднений такое представление как бы не вызывает.

Онтологические приложения комплексного теста

Онтологический контекст проблематики электронной культуры задает необходимость вписывания концепта экзистенции Я в картину высокотехнологичного мира. В конвергентное развитие и функционирование НБИКС-технологий надо включить мое сознание с невероятно сложным комплексом мыслей, переживаний, болей, намерений и прочего. Традиционная «проблема сознания» расширяется, она становится не психофизиологической, а психотехнологической проблемой. Ведь в контексте теста Тьюринга вполне корректно поднимается вопрос о реализации когнитивных феноменов посредством компьютерной электронной технологии: имитации для частичной замены человека в социокультурных сферах; моделирования – для изучения индивидуальных и общественных форм сознания; репродуцирования – для воспроизводства когнитивных функций на искусственных системах; креацирования – для проявления того, чего нет в природе. Для решения новой онтологической проблемы сознания, для вписывания Я и другого в контекст компьютерного мира

адекватным представляется постнеклассический функционализм. Это – не традиционный компьютерный функционализм, иначе называемый компьютеризмом, который совмещает функционализм Х. Патнэма («машинный функционализм», для которого характерен эгологический тезис «Я – машина Тьюринга») с контентным функционализмом его ученика – Дж. Фодора. Комплексный тест Тьюринга позволяет сформулировать новую форму функционализма, которая способствует смысловому конституированию компьютерного мира. Постнеклассический функционализм мы назвали *тестовым функционализмом* (функционализмом теста Тьюринга). В разнообразии современных функционализмов он в наибольшей степени идентичен машинному функционализму, восполняя то, что было упущено Х. Патнэмом: не был учтен принцип удостоверения когнитивной компетенции систем, не была обозначена роль тьюрингова судьи. В тестовом функционализме Я снабжаюсь концептуальным инструментарием изучения компьютерного мира – комплексным тестом Тьюринга. Возможна программно-техническая реализация теста. Это – не существенно, здесь нет заявки на премию Лойбнера. Главное – рациональное мировоззренческое позиционирование субъекта компьютерного мира к феноменам компьютеризованной реальности.

Праксеологическое приложение комплексного теста

Комплексный тест Тьюринга особенно интересен в праксеологическом контексте изучения электронной культуры. Возможна ли свобода в условиях запрограммированной реальности? В последнее время проблема свободы воли, свободы выбора в этих условиях часто звучит на конференциях по когнитивной тематике. Комплексный тест Тьюринга, на мой взгляд, инструментально способствует ее изучению. Исследуются альтер- и интерсубъективные отношения Я и другого посредством критического анализа атрибуций (приписываний) мышления, переживания, творчества, сознания, самосознания, любви и других х-когнитивных феноменов у-системам. Осуществляется постнеклассическое позиционирование: Я – главный субъект электронной культуры, активно конституирующий её альтер- и интерсубъективные смысловые пространства. Я освобождается от алгоритмической зависимости, осуществляя своеобразное компьютерное «эпохе»: при компьютерной репрезентации когнитивного феномена в формате функционалистской дескрипции Я включает эту программу в «базу знаний» комплексного тестирования электронной культуры. Данная программа принадлежит Я, она полно и точно, как

утверждают разные функционализмы, характеризует мое сознание и сознание другого. Но это – не Я и не другой. Поэтому такую программную репрезентацию моих когнитивных способностей надо по-гуссерлевски отбросить за «скобки». Если комплексно отбрасывать атрибуты, исключается очень многое. Однако самое важное и приватное сохраняется «в остатке». Остается все то, с чем не справляется современный функционализм: самость, квалиа, интенция, проекция «каково быть собою или кем-то», содержание опыта и многое другое, все то, что Д. И. Дубровский обозначает термином «качество субъективной реальности».

В машинном и тестовом функционализмах роли Я и компьютера радикально различны. Если в первом утверждается: «Я – компьютер», то во втором, напротив, «Я – не компьютер». Тестовый компьютеризм способствует программированию когнитивных феноменов, развивая компьютерный базис электронной культуры. Однако в большей мере он предназначен для решения собственно человеческих, хотя и компьютерно модифицированных, проблем личности, свободы, смысла жизни, бессмертия, социального идеала.

Позвольте предложить онтологическую формулу тестового компьютеризма: когнитивный феномен электронной культуры конституируется атрибутами компьютерно непредставимого Я, исполняющего функции комплексного теста Тьюринга. Тогда онтологический тезис таков: *non computato, ergo sum* – не компьютеруем, следовательно, существую. Он принципиально отличается от тезиса, который проявляется в рамках репрезентативного контекста изучения электронной культуры: *e-re-praesento, ergo sum* (если Я представлен в компьютерной технологии, следовательно, существую).

Эти онтологические тезисы характеризуют фундаментальные отношения между человеком и компьютерным миром. Их постановка инспирирована идеей комплексного теста Тьюринга. Поэтому мне кажется, что такой тест является категориальной концепцией философских исследований искусственного интеллекта, а философия искусственного интеллекта способна выполнять многие функции классической философии, но в современном контексте, в контексте электронных технологий компьютерного мира, в который «заброшен» сегодняшний человек.

Заключение: прохождение теста Тьюринга как национальный проект

Гонка программ в прохождении теста Тьюринга сегодня стала сродни национальной идее первенства в построении искусственного интеллекта. Страна обогнала другие страны, если ее программа впервые прошла тест Тьюринга. Это очень напоминает космическую гонку времен холодной войны.

На нашей конференции выступали представители робототехнического центра «Сколково» с ошеломляющим сообщением о том, что российская программа впервые в 2014 году прошла тест Тьюринга. Однако Дорандо Ду, вице-президент мощной китайской компании распознавания речи IFLATECH, полагает, что впервые именно китайские программисты прошли тест Тьюринга. Это было в 2016 году, как бы в честь 60-летия искусственного интеллекта. Про российскую программу китайцы не знают или умалчивают. Но и наша программа вряд ли была первой. Так, американец Марк Хамприус полагает, что первой была его программа Mgonz. Она тестировалась на Лойбнеровских соревнованиях в 1995 году и уже тогда была чатом реального времени, в неё были заложены элементы неожиданности (ненормативная лексика), и она работала в Интернете. Но и до Mgonz было разработано много подобных чатов [10, с.45-46; с.47-48]. А как быть со знаменитой «Элизой» американско-германского ученого Джозефа Вайценбаума? Для секретаря Дж. Вайценбаума, которая любила «поболтать» с компьютером, эта программа уже пятьдесят лет назад, в 1966 году, прошла тест Тьюринга.

Возникает классический, но компьютерно модифицированный вопрос: а судьи (тьюринговские) кто?

ЛИТЕРАТУРА

1. *Turing, A.* (1950), Computing Machinery and Intelligence, *Mind* 59(236), pp. 433–460.
2. *Алан М. Тьюринг.* Может ли машина мыслить? (Под ред. Б. В. Бирюкова). М., 1960. С. 19-158.
3. *Алексеев А. Ю.* Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга? // Искусственный интеллект: междисциплинарный подход. Под ред. Д. И. Дубровского и В. А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.223-242
4. Тест Тьюринга: философские интерпретации и практические реализации: Материалы научно-практической конференции, посвящённой 60-летию публикации статьи Алана Тьюринга «Computing Machinery and Intelligence»; Под общ.ред. А. А. Костиковой, редакция и составление

- Н. Ю. Ключевой – М.: Алькор Паблишерс, 2011. – 136 с. – (Серия кафедры философии языка и коммуникации).
5. Алексеев А. Ю. Роль комплексного теста Тьюринга в современной философии искусственного интеллекта// 10 лет – Научному совету РАН по методологии искусственного интеллекта. Материалы симпозиума, 26 марта 2015 г., Институт философии РАН, г. Москва / Ред.: Алексеев А. Ю., Дубровский Д. И., Лекторский В. А. – М.: ИИНтелл, 2016. – 142 с. – (Сер.: Философия искусственного интеллекта) – С. 118-127
 7. Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. – М.: ИИНтелл, 2013 г. – 304 с.
 8. Saygin, A. P., Cicekli, I. & Akman V. 2000. Turing test: 50 years later. *Minds and Machines* 10:463-518.
 9. Epstein, R., Roberts, G., Beber G. 2008. Parsing the Turing Test Philosophical and Methodological Issues in the Quest for the Thinking Computer. Springer Science + Business Media B.V. – 517 p.
 10. *Методологические и теоретические аспекты искусственного интеллекта*. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А. Ю. Алексеева – М.: МИЭМ, 2006. – 192 с.,
 11. Макаров В. Л. Получение нового знания методом компьютерного моделирования// Искусственный интеллект: междисциплинарный подход. Под ред. Д. И. Дубровского и В. А. Лекторского – М.: ИИНтелЛ, 2006. – 448 с. – С.5-11.
 12. Putnam, H. (1960). *Minds and machines*. In (S. Hook, ed.) *Dimensions of Mind*. New York University Press. Reprinted in *Mind, Language, and Reality* (Cambridge University Press, 1975).
 13. Putnam, H. (1967). *The mental life of some machines*. In (H. Castaneda, ed.) *Intentionality, Minds and Perception*. Wayne State University Press. Reprinted in *Mind, Language, and Reality* (Cambridge University Press, 1975).
 14. Putnam, H. (1981). *Reason, Truth, and History*, Cambridge: Cambridge University Press. – 222 p.
 15. Putnam, H. (1997). "Functionalism: Cognitive Science or Science Fiction?" In D. M. Johnson and C. E. Erneling (eds.), *The Future of the Cognitive Revolution*. Oxford: Oxford University Press, pp. 32–44.
 16. Лекторский В. А. Философия, искусственный интеллект и когнитивная наука// Искусственный интеллект: междисциплинарный подход. Под ред. Д. И. Дубровского и В. А. Лекторского – М.: ИИНтелЛ, 2006. – 448 с. – С.12- 21.
 17. Levin, Janet (2013), «Functionalism», *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/functionalism/>> .
 18. Horst, Steven (2011), «The Computational Theory of Mind», *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>>.

Я-МИРОВОЗЗРЕНИЕ: ПРОЕКТ МОДИФИКАЦИИ СТРУКТУРЫ ВНУТРЕННЕГО МИРА

I-WORLDVIEW: A PROJECT TO MODIFY THE STRUCTURE OF THE INNER WORLD

库瓦琴 «我是世界观：心理结构的改造工程»

Кувакин

Валерий Александрович

доктор философских наук, заслуженный
профессор Московского университета,
профессор кафедры истории русской философии
философского факультета МГУ имени
М. В. Ломоносова, г. Москва

Kuvakin Valery A.

Doctor of Philosophy, Honored Professor of the
Moscow University, Professor of the Department
of the History of Russian Philosophy, Faculty of
Philosophy, M. V. Lomonosov Moscow State University,
Moscow



Аннотация. Поднимается проблема программирования Я в интеллектуальных системах. Изучаются фундаментальные контексты компьютерной модели «Я»: бытие, мировоззрение, Я-воззрение. Анализируется пространство субъективной реальности с целью отыскания новых ресурсов интеллектуальной и психологической безопасности человека в когнитивных технологиях.

Ключевые слова: Я, бытие, мировоззрение, Я-мировоззрение, программирование «Я».

Annotation. The problem of programming I in intellectual systems is raised. The fundamental contexts of the computer model «I» are studied: being, worldview, I-view. The space of sub-reality with the purpose of finding new resources of intellectual and psychological security of a person in cognitive technologies is analyzed.

Keywords. I, being, worldview, I-worldview, programming «I».

В предлагаемом материале анализируется пространство внутреннего мира с целью отыскания новых ресурсов интеллектуальной и психологической безопасности человека в условиях беспрецедентного давления на него информации нормативного, перформативного, суггестивного и манипулятивного характера.

Предварительное замечание. Коль скоро речь идет о Я, то здесь предлагается такое его понимание: Я – это всё вместе. Я – это бытие, это ничто, это неизвестно что. Высказанные утверждения можно обсуждать, но в любом случае Я – это не пустая оболочка, безразмерная точка, вымысел или мысль, которой ничего не соответствует. Прежде всего, Я – это ресурс, это возможность вершинности и завершения самосозидания личности, с которого начинается полноценная равноправная по отношению к миру жизнь человека. В Я человек вновь и вновь обретает свое собственное существование, как бы вновь рождается и подтверждает себя. Главная функция Я – осуществление процедуры самоидентификации, самоопознания человека, его «почти» или «уже» рождение и саморождение. Я-воззрение – это система суждений об автономии и равноправии человека перед лицом мира.

О том, что Я не является фикцией, комплексом идей или мировоззрением, говорят, в частности, вполне материальные структуры мозга, которые отвечают за самоидентификацию человека. Я тесно связано с самосознанием и использует его как главный инструмент своего функционирования. В этой связи один из крупнейших ученых в области когнитивной нейронауки Вилейанур С. Рамачандран отмечает: «Самосознание. Этот аспект «я» почти очевиден; «я», которое себя не осознает, оксюморон... самосознание, возможно, частично зависит от рекурсивного использования мозгом зеркальных нейронов, что позволяет нам видеть самих себя со стороны (аллоцентрически)» [1].¹

Это предварительное замечание позволяет перейти к основной проблеме статьи: можем ли мы изменить структуру внутреннего мира: Я – самосознание – сознание? Сначала кажется, что нет. Но мы можем так просветить ее нашим сознанием и самопознанием, что будем в состоянии увидеть то, что редко удастся заметить: границу между Я и самосознанием. Если же ее нет, то мы можем провести ее и установить демаркацию между Я и самосознанием. Тем самым мы покушаемся на структуру внутреннего мира. Что это за операция?

¹ Кроме указанного, Рамачандран называет такие признаки Я, как целостность, постоянство, пребывание в теле, личность, социальность и свободную волю [1].

Что она значит? Зачем это нужно? Об этом – весь последующий разговор.

Я окружено и погружено в самосознание, но Я и самосознание не тождественны. Более того, здесь постулируется то, что на их границе есть нечто, что мы условно можем назвать складкой. Она складка потому, что там ничего нет, а если что-то и залетает туда, то долго там не удерживается, а точнее говоря, туда и залететь ничего не может. Положим, эта складка – некое схлопнутое виртуальное пространство между Я и самосознанием, пространство как бы сложенное или «сдутое» и обычно незаметное для самого человека. Но эту складку можно «разгладить» нашим сознанием, устремленным в глубь человека. Если ее «разгладить», то во вновь образовавшееся пространство Я может заложить идеи, которые будут образовывать буферную зону, «подушку безопасности» (как в автомобиле), которая может предохранить Я от непосредственного соприкосновения с контентом (содержанием), окружающим это Я. (Строго говоря, подобных демаркаций мы можем провести во внутреннем мире сколько угодно и где угодно, вопрос лишь в том, можем ли мы при этом создавать на базе таких демаркаций (пограничных полос) качественно новые контенты, обладающие принципиально новыми функциями.)

Эту процедуру удобнее представить не столько как изменение структуры внутреннего мира, сколько как её восстановление или восполнение, как если бы мы, надувая резиновую игрушку заведомо данной формы, вдруг обнаружили некую неизвестную прежде складку, наполняемую воздухом, отчего форма этого предмета изменилась и стала нормальной, такой, какой эта форма в нашем случае и должна быть у развитого внутреннего мира или просто личности. Таким образом, из складки мы вначале делаем некоторое новое пространство, разглаживая нашим сознанием то, что было сложено, и далее с его же помощью наполняем эту зону определенным содержанием.

Какие же идеи и суждения может высказать Я о самом себе? В чем их особенности? Каково их значение для человека? Специфика этого нового контента (содержания зоны безопасности) состоит, во-первых, в том, что это идеи Я о самом себе. При этом Я не нагружено и не обременено никаким мировоззрением, убеждениями или идеологией, оно в этом смысле чистое и беспримесное Я. Во-вторых, эти идеи хороши и полезны потому, что позволяют нашему Я наладить здоровые отношения между ним и контентом (содержанием) самосознания и сознания. Эти отношения становятся опосредованными,

подобно разговору, происходящему между людьми, находящимися на приличном расстоянии друг от друга, – его и образует подушка или зона безопасности, зазор, пограничная полоса или дистанция свободы. В-третьих, сами эти идеи не являются *миро*-воззрением, потому что Я видит и высказывает идеи не о мире, а о себе, т.е., повторю, это не *миро*-воззрение, а воззрение Я на самого себя, в точном смысле *Я-воззрение*. Ничего внешнего и потенциально опасного для Я здесь нет. В-четвертых, с этими Я-идеями не конфликтует ни одно из известных мировоззрений. С суждениями Я о Я уживается практически любое мировоззрение, за исключением очевидно деструктивных: террористических, криминальных и т.п.

Я-воззрение открыто любому мировоззрению и никоим образом не ограничивает ни свободы мышления, ни свободы совести человека. За ним можно признать статус *транс*-воззрения или *мета*-мировоззрения, поскольку оно идет поверх каждого из них в отдельности, не будучи ни одним из них. Но его открытость любому мировоззрению не исключает его несовместимость с такими воззрениями, которые можно отнести к разряду «анти-Я-воззрения». Это странное словосочетание означает лишь то, что есть воззрения, которые прямо враждебны нашему Я и предполагают уничтожение человека вместе с его Я или исчезновение Я в социуме, нирване, ничто, в океаническом разуме (как он описан, например, в «Солярисе» С. Лема), в Матрице или в чем-то другом. О том, каким причудливым путем происходило осмысление Я в мировой философии, см [2].

Еще один тип «анти-Я-воззрения», которое не может совеститься с идеями я-воззрения, – это такое воззрение на человека, которое пронизано ненавистью к человеку и – с неизбежностью – к его воззрениям, к самому Я этого человека. Таких людей – мизантропов – мало, но они есть. Это люди, которые ненавидят не только других, но и себя, и не учатся любить и уважать ни себя, ни других. Между тем, более серьезной опасностью для человека является сама возможность каждого из нас быть доведенным до такого состояния, из которого вырастает ненависть к самому себе.

Каково же содержание этого *Я-воззрения*? Какие суждения наше Я может сформулировать относительно самого себя? Вот минимум суждений Я о себе самом.

Я есмь, я существую.

Я – это Я и ничего больше, и ничего меньше.

Я приоритетно и ближе всего к себе по сравнению с тем, что окружает меня во внутреннем мире человека и за его пределами.

Я несопоставимо более высокая ценность, чем любая идея или мировоззрение, потому что живу и должен жить Я, хотя и с помощью идей, всего контента моего внутреннего мира.

Я не допускаю, чтобы идеи жили за счет меня, моей жизни. Я должен создавать автобиографию, а не находящееся вокруг меня мировоззрение должно писать биографию своей жизни во мне и за счет моей жизни.

Я самосохраняюсь, не отдаюсь никому и ничему, ни на что не обмениваюсь, тем не менее я помогаю человеку стать и быть личностью.

Я сохраняю вокруг себя пространство свободы; это необходимо мне для маневра и динамики, для управления и координации всего остального в человеке.

Я оперирую не одним, а множеством мировоззренческих парадигм, устанавливая в этой сфере сознания плюрализм и партнёрские отношения меня с ними при сохранении моего приоритета и права вето: в конце концов, решаю Я, а не мировоззрение.

Я неотделимо от своей плоти, а не только от сознания, поскольку Я суть единство материального и виртуального и распределено как в сознании, так и в теле. В этом смысле Я вездесуще и мне до всего есть дело.

Хотя Я и уникально, единично и сингулярно, у меня родовая природа, что делает для меня реальным единство планетарного человечества.

Я как таковое не имею ни расы, ни пола, ни возраста, ни какого-либо социального статуса. В физическом и биологическом смысле Я локализовано в индивиде и неотделимо от него, но виртуально Я нахожусь вне вселенского пространства и времени.

Как легко убедиться, почти все эти суждения вполне тривиальны, в принципе понятны. Они не предполагают ничего внешнего, привнесенного в Я извне. Число этих суждений, составляющих минимальный набор принципов Я-воззрения, скорее всего, может быть увеличено или даже уменьшено. Главное при этом – не вносить в Я-воззрение ничего не присущего самому этому Я, хотя «отблесков» мира в нем достаточно много. К одному из них относится ценностное суждение Я о самом себе и социальный подтекст большинства из этих суждений. Это означает, что наше Я, будучи чем-то особенным и «выпадающим» из мира, не вмещающимся в него, тем не менее, примеряет

к себе параметры, которые присущи не только Я, но и к другим вне его лежащим реальностям, т.е. природе и обществу¹. Наиболее важные из этих параметров – неизвестность (загадочность, сокрытость, тайна человеческого Я даже для самого себя), существование, действительность (оно действительно), свобода, независимость, автономия, приоритет, ценность, внутренний диалог, партнерство, безопасность, решимость. К этому следует добавить такое, казалось бы, далекое от науки и философии понятие, как «простор». Возможно, простор лишь одна из модификаций свободы и динамики, но я бы выделил его специально, поскольку именно простор, безграничность и неопределенность нашего Я, обрамление его свободой и автономией Я-воззрения суть необходимейшие условия его существования, жизнепроявления и жизнотворчества.

Я-воззрение и изложенный выше мини-кодекс Я, выражающий его декларацию независимости и статус, представляется вполне очевидным, даже банальным, хотя «добраться» до этой банальности философским и ретроспективным методом достаточно сложно (лично у меня это заняло многие годы). Однако то, что сложно сделать философски, по-своему и довольно прозрачно формулируется языком права (см. ст. 28 Конституции РФ или ст. 18 и 19 Всеобщей декларации прав человека)².

Какова работа Я-воззрения, что оно делает или что должно делать для нас? Непосредственно, не будучи активированным, «задетым», разбуженным, оно остается «закулисным». Но бодрствуя,

¹ Как пишет В.С.Рамачандран, ««Я» поддерживает преувеличенное чувство личностности и автономии, которое выдаёт её тесную связь с умами других людей. Случайно ли, что почти все наши эмоции обретают смысл лишь в отношении других людей? Гордость, высокомерие, тщеславие, амбициозность, любовь, страх, милосердие, ревность, гнев, надменность, гуманность, жалость, даже жалость к себе – ни одно из этих чувств не имело бы никакого смысла в социальном вакууме» [1].

² «Каждому гарантируется свобода совести, свобода вероисповедания, включая право исповедовать индивидуально или совместно с другими любую религию или не исповедовать никакой, свободно выбирать, иметь и распространять религиозные и иные убеждения и действовать в соответствии с ними» (Конституция РФ, ст. 28. «Каждый человек имеет право на свободу мысли, совести и религии; это право включает свободу менять свою религию или убеждения и свободу исповедовать свою религию или убеждения как единолично, так и сообщая с другими, публичным или частным порядком в учении, богослужении и выполнении религиозных и ритуальных обрядов». «Каждый человек имеет право на свободу убеждений и на свободное выражение их; это право включает свободу беспрепятственно придерживаться своих убеждений и свободу искать, получать и распространять информацию и идеи любыми средствами и независимо от государственных границ». Всеобщая декларация прав человека. Ст. 18 и 19).

питаюсь энергией самосознания, думая о самом себе, осмысляя себя, Я «взрослеет», создаёт суждения и смыслы о себе и, далее, – о своем ближайшем окружении: о мировоззрении, о содержащейся в сознании информации, знаниях, опыте и т.д. Такому активному Я легче помочь человеку обрести бóльшую жизненность, адаптивность, конкурентоспособность, креативность, помочь быть мировоззренчески здоровым существом.

Но насколько эффективна эта буферная зона Я-воззрения как метамировоззрения и службы внутренней безопасности перед виртуальностью Интернета, тотальных масс-медиа и коммуникаций? Она зависит, грубо говоря, от двух факторов: (1) достаточной осведомленности человека о своем внутреннем мире, способности усвоить идею Я-воззрения и обрести его; (2) от наличия хотя бы минимального уровня демократии, позволяющей реализовать дистанцию свободы (зазора между убеждениями/идеологиями и его собственным личностным началом), как в целом и остальные свои права человека и гражданина.

Мы знаем теперь, что между Я и мировоззрением образуется (может образоваться) «предбанник», зона перехода, буфер, зона безопасности, состоящая из суждений Я о Я, не включающих суждения о чем-либо ином, не являющемся этим Я. То есть среди этих суждений нет явных суждений о не Я.

В теоретическом плане нашему Я присуща наивность и оголённость райского человека. Если сдуть Я-воззрение как подушку безопасности, то Я и, следовательно, индивида легко одеть любыми одедами (мировоззрениями и идеологиями) – от царских до нищенских. Значит ли это, что наше Я, в конце концов, нечто одинаково сильное и слабое? Да, видимо, это так. Поэтому без самоутверждения, самоконституирования, без нормативных, его собственных идущих изнутри защитных процедур здесь не обойтись. Их-то и предлагает современный светский, научный, нерелигиозный гуманизм, наиболее бескорыстно защищающий человека и, следовательно, его Я.

Почему я упомянул нерелигиозность? Причина проста. Наше Я естественно. Оно ничуть не религиозно не потому, что атеистично и испытывает какой-то изначальный негативизм к трансцендентному, а потому, что ему в принципе противопоказано отказываться от собственного приоритета по отношению к себе¹. Более того, насколько

¹ Не случайно в одном из документов нерелигиозного, светского гуманизма отмечается: «Гуманизм провозглашает достоинство человека, его уникальную роль во Вселенной, его ценность, приоритетную и безусловную по отношению к себе, но относительную и равноправную по отношению к природе и обществу» [7].

Я нерелигиозно, настолько же оно и внемирно, т.е. не растворимо в чем-то внеположном ему самому, независимо от того, реальность ли это или что-то вымышленное. Парадоксально можно сказать, что Я от мира, но вне его. Возможно, при рождении Я происходит какой-то метафизический отскок (но не разрыв) Я от мира. В этом состоит одна из особенностей сингулярности, уникальности и нередуцируемости человека.

Теперь, оснащенные Я-воззрением, начнем оглядываться внутри себя. Кому принадлежит внутренний мир? По форме, как структура – мне, т.е. Я (самости, эго и т.п.), по праву единородства и связи психического и физического, ментального и телесного во мне как человеке. Как виртуальная структура внутренний мир принадлежит мне так же, как мое тело – моему телу, хотя в то же время последнее явно принадлежит, точнее, причастно миру – ведь мы сделаны из того же материала, что и космос. Но из чего сделано наше Я – сказать трудно, во всяком случае, оно, будучи разбуженным, явленным человеку изнутри, сразу дает понять, даже почувствовать, что оно, выглядывая из неизвестности и оглядываясь кругом, «высовывается» из мира, переживая себя одновременно и совместимым и несовместимым во все и вся. При этом, диссипатируя, оно передает это переживание всему человеку, вплоть до изменения его эмоционального и телесного состояния. Чувство Я говорит, что есть в каждом из нас что-то особое, хотя и рожденное в мире, но как бы не от мира сего. Оно – полупрозрачно, таит в себе какую-то загадку и бесконечную глубину. Мы заглядываем в него (точнее, само оно заглядывает в себя с помощью сознания и всех других познавательных инструментов), но вывернуть его наизнанку (иначе: Я вывернуться из самого себя) мы не можем, да и едва ли, если разумны, желаем этого. Оно – наш гарант и инструмент идентичности, нашей подлинности и уникальности. Это не означает, что Я твердо знает, что не может этого сделать – вывернуться наизнанку. Нам неизвестны возможности Я. Но можно предположить, что попытка человека «взломать» свое Я и внедриться в него может быть пагубной, как пагубна игра ребенка с боевой гранатой. А то, что человечество сегодня не более чем подросток, при трезвом размышлении может быть понятно любому образованному человеку.

Наше Я пребывает на границе нашего существа (человека как целостности) и мира, но обладая способностью к трансцендированию, прорыву, выходу из себя, пульсируя на границе, выходя за неё, оно остается в её пределах (здесь мы имеем эффект раздвигающегося

горизонта). Тем самым Я выражает свою особенность, свою таинственность и неуловимость, которыми оно, несомненно, обладает как в отношении себя, так и в отношении всей тотальности человека, не говоря уже об окружающей действительности. О чем-то похожем говорит Карл Ясперс в своем учении о всеобъемлющем – *Umgreifende*.

Такова, в общих чертах, ситуация в отношениях Я и структуры внутреннего мира.

Хотя мир нам дан, размещен во внутреннем мире в формах ничто, т.е. сознания, виртуально, это не значит, что мы остаемся ни с чем. Человек устроен таким удивительным образом, что мы это отражение можем трансформировать в реальное действие через мышление, переживание, созидая или трансформируя бытие в мире через практику и творчество. Строго говоря, мы усваиваем из окружающей нас тотальности мира как триединства действительностей не только (1) бытие (материальный, предметный, вещный, физический мир), (2) ничто (скажем, отсутствие чего-либо или виртуальные содержания Интернета, или история как память о бывшем, или символизм и «искусственность» культуры) и (3) неизвестность. Последнюю мы особенно ярко опознаем в психологических и экзистенциальных формах: в ситуациях уединённости или сокрытости, обрушившегося на тебя горя или счастья, любви или созерцания прекрасного и т. д. Тем более очевиден для нас «гносеологический» контакт с неизвестностью, когда мы устремляемся к ней, подвластные и одержимые её загадочностью, смутными ожиданиями новизны, открытия, прорыва в неизвестность здесь-и-теперь, заведомо оснащенными жадной знания – этим тралом, забрасываемым в неизвестность.

И в заключение снова спросим: может ли человек менять содержания внутреннего мира? Очевидно, что да: информационное взаимодействие разума человека и мира – это невероятно мощный двусторонний изменчивый процесс. Может ли человек менять структуру внутреннего мира? Да. Об этом говорится здесь в статусе гипотезы, излагаемой в качестве предлагающего объяснения. Если говорить о самосознании и Я, то мы можем быть глухи к ним, не развиты ни телесно, ни психически, ни культурно настолько, что самосознание и центрирующее его Я просто схлопываются, точнее не раскрываются, оставаясь пустыми виртуальными пространствами, не дающими знать о себе. Они не исчезают, а как бы засыпают, летаргируют. Что касается Я, то оно вообще редкий гость у себя дома, обычно проявляя себя через что-то иное. Но случается, что человек вдруг или

постепенно доходит до такого состояния, которое дает ему почувствовать, что «задето его Я». И не только задето, но и возмущено, унижено, оскорблено, или, напротив, признано, удовлетворено и т.д. Однако эти разные состояния нашего самосознания и Я не означают, что мы меняем их архитектуру или структуру. Это значит, что мы просто активируем их или нет.

Вместе с тем, вновь подчеркнем, между Я и самосознанием есть граница, которая может быть изменена более радикальным образом. Это то, что выше было названо «складкой», превращаемой в дистанцию свободы, заполняемой Я-суждениями, образующими Я-воззрение. Возможно, речь идет о банальности, которую я возвожу в ранг чуть ли не открытия, но именно эта особая сфера или «пустое место», насколько мне известно, нигде не описана и не осмыслена. Повторю еще раз: возможно, то, что я разумею под складкой и всем с нею связанным, уже давно обнаружено, понято и проинтерпретировано, но просто выражено иным языком и в иной парадигме философской антропологии и психологии личности.

Иногда говорят об открытии человеком своего Я, иногда говорят о человеке высокого и интенсивного уровня рефлексии. Во всяком случае, изучение внутреннего мира и самопознание можно представить себе как путешествие, полное научных и личных открытий. Между тем в области психологии и философии немало ученых, которые де факто не признают, по меньшей мере, двух вещей: существования Я как чего-то отличающегося от мировоззрения этого Я. Во-вторых, само Я склонны понимать как Я-концепт и только¹. То есть устанавливается либо знак тождества между Я и мировоззрением, либо Я редуцируется к представлению о нем самого человека или к содержаниям окружающей его действительности в ходе ее отражения и интериоризации. При этом латентно предполагается, что какого-то трансцендентального (родового) Я не существует, проще говоря, нам нечего сказать о каких-то стационарных и даже внеличностных характеристиках этого феномена. Оно тем самым релятивируется и субъективируется в психологическом смысле этого слова. Фактически оно сводится к понятию личности или индивидуальности.

В целом, по наблюдениям А.Л.Доброхотова, «XX век дает нам четкое размежевание этих направлений. Ряд направлений все-таки пытается обновить интуицию чистого «Я»: феноменология Гуссерля,

¹ Одним из первых в современной культуре единство Я разрушил З. Фрейд, предложив свою инструментальную, но научно так и не доказанную, гипотезу о нескольких Эго.

экзистенциализм, персонализм, новые версии психоанализа, герменевтика. А другой ряд продолжает традицию растворения «Я» в эмпирических, исторических или психологических стихиях» [2]¹.

Часто, а на обыденном уровне практически всегда, наше Я отождествляется с мировоззрением. В результате мы не только не ставим вопрос о буферной зоне или подушке безопасности для Я, но и отдаем его на потребу мировоззрениям и мнениям о себе, т.е., в конце концов, суждений других людей о нашем Я.

Между тем, все более очевидной становится необходимость конституирования Я-воззрения наряду с Мировоззрением человека. Такой тандем, приоритет, ведущая роль в котором принадлежит первому, способен помочь нам, людям, обрести свой собственный полноценный статус как четвертой субстанциональной действительности наряду с бытием, ничто и неизвестностью [6].

В итоге речь идет о переопределении личности. В свете сказанного, аутентичная личность – это Я, оснащенное двумя воззрениями: (1) эксцентричный комплекс, состоящий из собственно Я, Я-воззрения и (2) миро-воззрения. В целом он эксцентричен потому, что его экзистенциальный и ценностно-смысловой центр смещен к Я, как субъекту этих воззрений. Это Я-центрированный комплекс. При этом Я-воззрение образует систему внутренней безопасности Я и далее самой личности, а Мировоззрение позволяет ей вписаться в объективный мир (включающий телесность человека), адаптироваться к нему, жить и действовать в нем. Можно сказать, что такая личность – это человек с усиленным центром тяжести и мощной фоновой рефлексией, повышающими его жизнестойкость, самообладание, свободу, решимость и ответственность. Сколь ни важны генетические и объективные: природные, социальные, культурные, – детерминации в становлении и рождении человека, в открытии и пробуждении им своего Я, решающую роль могут и должны играть факторы самодетерминации, принцип приоритета личности в выборе и решении, в понимании и истолковании окружающей действительности и самого себя.

Каким образом предлагаемая концепция уточняет и углубляет определение и самоопределение человека? Посредством Я-воззрения мы имеем возможность дать более полную психологическую, философскую, аксиологическую и экзистенциальную экспозицию специфически человеческого присутствия, действия и жизни в мире. Она дает метафизическое обоснование той статусной составляющей человека,

¹ Что касается растворения в «психических стихиях», см. [8].

которая относится только к нему и отсылает к собственно человеческому в нем. Я-воззрение доводит обоснование суверенной ценности и достоинства человека до его логического конца, подводя под это обоснование аргумент от метафизики. Окружающие Я Я-утверждения подобны сторожевым башням, стерегущим действительность Я как гаранта свободного и автономного существования человека. Они составляют Декларацию его действительности и содержательно подтверждают его самоидентичность.

К существующим юридическим, моральным, экономическим и экологическим обоснованиям достоинства, прав и свобод индивида Я-воззрение добавляет целостное объяснение человеческого феномена как сингулярного и уникального единства бытия, ничто и неизвестности в бытии, небытии и неизвестности мира. Но главным для человека как человека вчера, сегодня и навсегда остается его безопасность – психологическая, интеллектуальная, мировоззренческая, идеологическая, информационная. Внутренняя безопасность человека – это предпосылка и основа его внешней безопасности, его гармоничных и безопасных отношений со своей плотью, своим внутренним миром, с себе подобными, обществом и природой.

И в заключение о важной прикладной стороне дела, возможности моделировать и программировать структуры Я-воззрения в рамках общей стратегии создания искусственного интеллекта. По моему глубокому убеждению, эта задача не представляется на порядок сложнее, чем другие модельные практики ИИ. Это следует уже из того, что на генетическом уровне сама проблема Я – это проблема самоидентификации, прямой и обратной связи в рамках фундаментального тождества Я с самим собой, $Я \equiv Я$. По мнению генетиков, программы самоидентификации и демаркации для живых организмов одни из самых древних и закладываются уже на первых стадиях формирования живого организма [9, с. 256–259], что обусловлено, в частности ключевой задачей любого живого существа – проведением демаркации между «своей» и «чужой». Это обстоятельство, с учетом прозрачности структуры Я-суждений, практически представляющих собой набор атомарных самодостаточных и самотождественных утверждений, позволяет с большой долей уверенности говорить о простоте исходных данных для программирования и моделирования феномена Я-воззрения.

По словам Д. И. Дубровского, с точки зрения нейронауки, «... наше Я представлено в мозге особой структурно-функциональной подсистемой, которую называют **Эго-системой головного мозга** (или

Самостью). Она включает генетический и биографический уровни диспозиционных свойств индивида, образует высший, личностный *уровень мозговой самоорганизации и управления*, который образует сознательно-бессознательный контур психических процессов. Именно на этом уровне совершаются те кодовые преобразования, которые представляют наше Я и информацию в «чистом» виде (т.е. в *качестве СР*) (субъективной реальности – В. К.), обеспечивают активность Я в форме *произвольных действий*, способность Я к *самоорганизации*, к поддержанию своей идентичности, реализации веровательных установок и целевых векторов. Эго-система воплощает **личностные** особенности индивида, способность личности к **волеизъявлению** [3, с. 52 – 53]. Высказанные соображения уже содержат в себе довольно содержательную и целостную стратегию моделирования Я-воззрения как вполне определенного виртуального паттерна внутреннего мира человека.

Как человек, в должной и возможной степени не озабоченный созданием полноценной службы внутренней безопасности своей личности, становится все более неадекватным реальностям окружающего мира, легкой добычей информационной реальности, так и ИИ, лишенный аналога Я-воззрения, не обладает всей полнотой своей безопасности и структурно-функциональной завершенности.

ЛИТЕРАТУРА

1. Вилейанур С. Рамачандран Мозг рассказывает. – М. 2012 // http://www.e-reading.mobi/bookreader.php/1024774/Ramachandran_-_Mozg_rasskazyvaet._Chto_delaet_nas_lyudmi.html/.
2. Доброхотов А. Л. Понятие «Я» в философии и культуре. 7 фактов о рождении философской концепции «Я» и ее трансформации в различных традициях и эпохах // <http://postnauka.ru/faq/34727/>
3. Дубровский Д. И. Проблема «сознание и мозг»: Теоретическое решение М., 2015.
4. Кувакин В. А. Стать собой. М., 2010.
5. Кувакин В. А. Скрытые зоны безопасности // Кувакин В. А. Личность и Просвещение: От постмодернизма к неомодернизму. М., 2011, с. 40 – 49.
6. Кувакин В. А. Человек в мире действительностей // Вестник МГУ, серия философия, 2005, № 3.
7. Манифест российских гуманистов // <http://www.humanism.ru/manifest2015.htm>
8. Маслоу А. Дальнейшие рубежи развития человека. М., 1999 // <http://www.golubinski.ru/socrates/maslow/index.html>
9. Karine A. Gibbs, Mark L. Urbanowski, E. Peter Greenberg. Genetic Determinants of Self Identity and Social Recognition in Bacteria

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НАУКА, МИФ, ФИЛОСОФИЯ

ARTIFICIAL INTELLIGENCE: SCIENCE, MYTH, PHILOSOPHY

备土露宁 «人工智能：科学，神话，哲学»

Петрунин Юрий Юрьевич

*доктор философских наук, профессор,
заведующий кафедрой математических методов
и информационных технологий в управлении
факультета государственного управления МГУ
имени М. В. Ломоносова, г. Москва*

Petrinin Yuri Yu

*Doctor of Philosophy, Professor, Head of the
Department of Mathematical Methods and
Information Technologies in Management of the
Faculty of Public Administration of M. V. Lomonosov
Moscow State University, Moscow*



Аннотация. В статье рассматриваются вопросы о внутренней противоречивости того явления, которое принято называть искусственным интеллектом, а также вопрос о том, какую роль способна играть философия в исследованиях по искусственному интеллекту.

Ключевые слова. искусственный интеллект, наука, миф, философия.

Annotation. The article analyzes the internal contradictions of the concept of artificial intelligence. The article also examines the question of what role can philosophy play in the research on artificial intelligence.

Keywords: Artificial intelligence, science, myth, philosophy.

Знаменательный юбилей исследований в области искусственного интеллекта, который отмечается в этом году (все-таки 60 лет!), – повод не только для торжественных речей, но и для осмысления пройденного пути, подведения итогов того, что достигнуто и в каком направлении нужно двигаться дальше. Я хотел бы выделить два вопроса, которые представляются мне важными. Первый касается определенной внутренней противоречивости того явления, которое принято называть искусственным интеллектом. Собственно говоря, само словосочетание «искусственный интеллект» явно парадоксально: как соединить, с одной стороны, слово «искусственный», «машинный», означающее нечто формальное, механическое, строго алгоритмическое, а, с другой стороны, слово «интеллект» – нечто креативное, оригинальное, не подчиняющееся никаким правилам? Второй вопрос – вопрос о том, какую роль играет (или может играть) философия в исследованиях по искусственному интеллекту?

Для правильного ответа на первый вопрос его целесообразно переформулировать следующим образом: можно ли свести прошедшие за последние десятилетия исследования по искусственному интеллекту к списку научных (и, возможно, инженерных) результатов. Что, собственно, достигнуто за довольно продолжительную историю этого научного направления?

Второй вопрос: искусственный интеллект – междисциплинарное направление науки, включающее в себя логику, психологию, лингвистику, информатику и т.п. Нужна ли здесь философия? Должен признаться, что словосочетание «философия искусственного интеллекта» звучит для меня как-то нарочито. Чем оно, собственно говоря, отличается от просто «искусственного интеллекта»? Попробуем кратко наметить правдоподобные ответы на эти два вопроса.

Несомненно, что термин «искусственный интеллект» прочно вошел во все современные языки. Но что он обозначает? Если провести обычное социологическое исследование, то мы обнаружим, что большинство респондентов прочно ассоциирует искусственный интеллект не с Джоном Маккарти, Марвином Минским или Лотфи Заде, а с Айзеком Азимовым, Арнольдом Шварценеггером в роли Терминатора или с «агентом Смитом» из «Матрицы». Одним словом, данное понятие, или, лучше сказать, образ формируется преимущественно не в научных исследованиях, а в жанре фантастики в литературе, кино, музыке, живописи, компьютерных играх. Как соотносятся научное и популярное представления об искусственном интеллекте?

Из истории известно, что литературные идеи часто опережают идеи научные в том смысле, что различные технические проекты, изобретения, предсказания развития человеческой цивилизации появляются раньше в искусстве, чем в строгой научной теории или тем более в реальном техническом воплощении. Художественное творчество служит как бы путеводной звездой для творчества научного. Вспомним хотя бы предвидения космических полетов, телефонов, телевидения и т.п. в произведениях писателей-фантастов. Однако в нашем случае, несмотря на некоторую внешнюю похожесть, существует принципиальное отличие.

Дело в том, что за шесть десятилетий развития научных исследований в области искусственного интеллекта, только об одном результате можно сказать вполне достоверно: «искусственный интеллект» до сих пор не создан. Да, ученые открыли много интересных методов, помогающих автоматически решать сложные задачи. Да, они научились моделировать многие важные аспекты человеческой интеллектуальной деятельности. Ученые даже создали машины и программы, успешнее и быстрее чем человек справляющиеся с традиционными задачами, например шахматной игрой или игрой го. Но полноценный искусственный интеллект, по общему признанию, не создан.

В этом и состоит принципиальное отличие ситуации. После создания телефона его использование в качестве элемента фабулы художественного произведения уже невозможно. Нельзя фантазировать о том, что уже существует. Об искусственном интеллекте – можно, потому что его нет и не ясно, когда он будет создан (если вообще будет). Возникает вопрос: не является ли сам искусственный интеллект всего лишь художественным вымыслом, неосуществимой мечтой человечества? Причем таким вымыслом, который увлекает, возбуждает даже неподготовленного в профессиональном (техническом) плане человека? Который связан с заветными чаяниями всего человечества?

Почти сорок лет назад было написано: «Высказываются опасения, что создать разумные машины – это значит впустить в наш дом полчища «завоевателей». На самом деле мы должны смотреть на себя как на осажденный гарнизон, который после десяти часов осады вдруг с облегчением замечает на горизонте пыль, поднятую спешащим нам на помощь отрядом» [1, с. 219].

А вот современные заголовки: «Илья Хель. Планы Маска на искусственный интеллект – больше, чем просто спасение мира» [2].

Или: «Алена Журавлева. Искусственный интеллект с правами бога» [3]. Это сообщается на порталах новостей в области науки и технологий, а не на литературных форумах!

Более того, вышеперечисленные формулировки – не просто метафоры. Исторические исследования показывают, что еще задолго до появления термина «искусственный интеллект» скрывающаяся за ним идея существовала не только в художественной, но и в мифологической форме (впрочем, современная гуманитарная наука доказала, что разница между последними формами условна и изменчива во времени).

Мифологема искусственного интеллекта связана с поисками утерянного адамова языка, сближавшего человека с Творцом. Его поиски привели к моделям универсальной обработки информации, открытию «информационной физики», созданию компьютеров (Подробнее о мифологеме искусственного интеллекта см.: [4; 5]). Умберто Эко писал, что «многие теории, ныне проверенные на практике, или практические достижения, по поводу которых мы теоретизируем (от систематики естественных наук до сравнительного языкознания, от формальных языков до проектов создания искусственного разума и новых горизонтов когнитивных наук), родились как побочные эффекты в ходе поисков совершенного языка» [6, с. 29].

Сакральный (мифологический) характер первоначальных поисков и сохранился сегодня в искусстве (и в массовом сознании): первые фильмы об искусственном разумном механизме были созданы ещё в начале прошлого столетия (*The Mechanical Statue and Ingenious Servant* (1907, dir. Stuart J. Blackman), *The Rubber Man* (1909, dir. Sigmund Lubin) и *Dr. Smith's Automaton* (1910). Бесконечные машинные войны в кино («Терминатор», «Бегущий по лезвию»), поражение человека созданной им техникой («Матрица») являются обратной, «неудачной» стороной попытки вернуться к блаженному адамову состоянию.

Неслучайно «экранизации» мифологемы искусственного интеллекта, как правило, связаны с апокалипсическими, или, говоря современным языком, алармическими видениями страшного будущего человечества, неявно напоминающими потерянную в благословенном прошлом главную цель существования. Поэтому, сколько бы ни производили дикувинок научные исследования в области искусственного интеллекта, общественное сознание ждет от них настоящих чудес,

способных решить самые важные вопросы, стоящие перед человеком и человечеством.

Обычно говорят, что искусственный интеллект – это моделирование интеллекта человеческого. При этом подразумевают, что человек-то интеллектом обладает априорно. Но вспомним: древнегреческие мудрецы сомневались в этом. Гераклит высказал эту мысль так: «Человеческая натура не обладает разумом, а божественная обладает» [7, с. 241]. «Человек.., – писал Платон, – это какая-то выдуманная игрушка бога» [8, с. 255]. В другом месте той же работы афинский мыслитель называет людей куклами, не способными к самостоятельным поступкам [8, с. 93]. Признавая разум мировой, универсальный, греки не могли принять, что его, говоря современным языком, можно смоделировать в отдельном автономном существе. В свое время предшественники нынешней версии искусственного интеллекта стремились моделировать не человеческий разум, а божественный, абсолютный. Вспомним, что логическая машина Раймонда Луллия совершенно неантропоморфна, а, скорее, космоморфна.

Что же из всего этого следует? Современное научное сознание склонно жестко разделять и противопоставлять миф и научную теорию, считая первую выдумкой, фантазией, искажением действительности, а вторую – более или менее точным слепком действительности. Конечно, это не соответствует истине. Об этом писали многие – не самые плохие философы от А. Ф. Лосева [9] и П. Фейерабенда [10] до К. Хьюбнера [11]. Не замечать того, что религиозное, мифологическое, художественно-образное начало в представлении об искусственном интеллекте неразрывно связано с рациональным, формально-логическим, значит не только быть пристрастным, необъективным, но и заранее обречь себя на неудачу в поиске решения столь заманчивой задачи.

Но, с другой стороны, как объединить два столь различных понимания того, что такое искусственный интеллект? Конечно, никакая наука, участвующая в столь грандиозном проекте, не способна к этому. Также никакой ненаучный – художественный, религиозный, мистический – подход не может сделать этого. Выйти за пределы узкопрофессиональных сфер познания, совместить два образа интеллекта, две стороны единой задачи призвана философия, которая включает в себя и рациональное начало, и внерациональное.

Именно философия способна постигнуть оксюморон искусственного интеллекта, снять неразрешимое противоречие задачи и направить совместные усилия различных дискурсов к единой цели.

ЛИТЕРАТУРА

1. Мичи Д., Джонстон Р. Компьютер-творец. М., Мир. 1987.
2. Илья Хель. Планы Маска на искусственный интеллект — больше, чем просто спасение мира. 22 декабря 2015. <http://hi-news.ru/technology/plany-maski-na-iskusstvennyj-intellekt-bolshe-chem-prosto-spasenie-mira.html>.
3. Алена Журавлева. Искусственный интеллект с правами бога. 29 января 2016. <http://spbit.su/news/n190046/>.
4. Петрунин Ю. Ю. Искусственный интеллект как феномен современной культуры // Вестник Московского университета. Серия 7. Философия. 1994, № 2, с. 28-34
5. *Призрак Царьграда. Неразрешимые задачи в русской и европейской культуре.* М., КДУ, 2006.
6. Эко У. Поиски священного языка в европейской культуре. Санкт-Петербург. ALEXANDRIA. 2007.
7. *Фрагменты ранних греческих философов.* Ч.1. От эпических теокосмогоний до возникновения атомистики. М., 1989.
8. Платон. Законы. // Соч. в четырех томах. Т.4. М., 1995.
9. Лосев А. Ф. Диалектика мифа. / Лосев А. Ф. Из ранних произведений. М., Правда. 1990.
10. Фейерабенд П. Против методологического принуждения. Очерк анархистской теории познания. / Пер. с англ. Благовещенск. 1998.
11. Хюбнер К. Истина мифа. / Пер. с нем. – М.: Республика, 1996.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СЕМИОТИЧЕСКАЯ ДИАГНОСТИКА АТТРАКТОРОВ КОНВЕРГЕНТНЫХ ТЕХНОЛОГИЙ

ARTIFICIAL INTELLIGENCE: SEMIOTIC DIAGNOSTICS OF CONVERGENT TECHNOLOGY ATTRACTORS

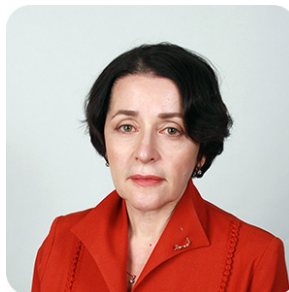
美理刻盖杂强 «人工智能：融合技术吸引
子会聚的诊断»

**Мелик-Гайказян Ирина
Вигеновна**

*доктор философских наук, профессор, зав.
кафедрой истории и философии науки, Томский
государственный педагогический университет,
г. Томск*

Melik-Gaykazyan Irina V.

*Professor, DSc in Philosophy, Head of the
department OF History and Philosophy of Science,
Tomsk State Pedagogical University, Tomsk*



Аннотация. «Искусственный интеллект» входит в систему конвергентных технологий, вызывающих нелинейную динамику всех компонент современной социокультурной действительности. Методы нелинейной динамики – это численные методы, поэтому применение релевантной методологии требует решения проблемы измерения в науках об обществе и человеке. Конвергентный характер технологий делает актуальной диагностику спектра аттракторов в динамике социокультурных систем. Паллиативным решением этих проблем являются процедуры семиотической диагностики, предлагаемые в статье. Связь «синергетика-семиотика» установлена на основе оригинальной трактовки феномена информации.

Ключевые слова: «семиотический аттрактор», семиотическая диагностика, биоэтика.

Annotation. «Artificial intelligence» is a part of the convergent technology system; these technologies cause the nonlinear dynamics of all contemporary socio-cultural reality components. Nonlinear dynamics methods are numerical methods, thereby the use of the relevant methodology requires the measurement problem solution in the society and human sciences. The convergent character of the technologies makes the attractor spectrum diagnostics in socio-cultural system dynamics relevant. The semiotic diagnostics procedures proposed in the paper are a palliative solution for these problems. The relation «synergetics-semiotics» has been determined on the base of the original interpretation of the information phenomenon.

Keywords: «semiotic attractor», semiotic diagnostics procedure, bioethics

Трактовки «искусственного интеллекта» имеют свою историю, сформированную конкуренцией исследовательских программ, опытом их технических реализаций, философских дискуссий вокруг содержания, направлений и перспектив этих программ. Вся многозначность и дискуссионность трактовок обладает областью пересечения, в которую входит признание того, что «искусственный интеллект» тесно связан с феноменом информации, воплощаемым в соответствующих технологиях. Их присутствие зафиксировано в аббревиатуре NBIC-технологий. Потенциалы этих технологий требуют специфического регулирования [1], поскольку они способны модифицировать человеческую природу [2], попадающую в локус их конвергентного схождения. Подобная потенциальность делает самоочевидной актуальность гуманитарной экспертизы целей NBIC-технологий. Проблему составляет отсутствие у гуманитарных наук точных методов диагностики этих целей и однозначного понимания природы информации. Взаимосвязь четырех оснований указывает путь решения проблемы:

1) символы растут [3], что стало квинтэссенцией процессуальной сущности семиозиса, открытой Ч. С. Пирсом;

2) устройство революций в символизме есть конечная мудрость управления [4], выражающее у А. Н. Уайтхеда применение его процессуальной онтологии события к пониманию сути воздействий символов;

3) событие как спонтанность, разделяющее прошлое и будущее, и формирующее аттракторы, которые меняют направления эволюции и темп устремления к этим объективным целям нелинейной динамики [5], что стало применением у И. Р. Пригожина философии

процесса А. Н. Уайтхеда в его программе исследований самоорганизации сложных открытых систем разной природы;

4) будущее временит настоящее [6], формулирующее позицию С. П. Курдюмова о структурах-аттракторах и процессуальности структур в его дискуссии с позицией И. Р. Пригожина.

Связь образует встречное движение естествознания и гуманитарного знания к трактовке того, в чем состоит общая сущность различных процессов – процессов символизации, процессов управления, процессов самоорганизации – и воздействию целевой причины на сценарии этих процессов. Для естествознания долгое время трудность представляла трактовка целевой причины без обращения к гипотезе о существовании «надприродных» сил, но язык математики избавлял от сложности различения процессов и функций состояния, например, на основе четкого различия свойств бесконечно малых величин – полных дифференциалов и частных производных. Для гуманитарного знания трактовки целевых причин составляют всю его традицию, но дарующая эту традицию вольность языка оставляет путаницу в представлениях о процессах, делая из слова «процесс» некую фигуру речи. Примером служит употребление слов «информационный процесс» при понимании самой информации как вещества, которое можно накопить или переработать; как функции состояния в рассуждениях о соотношении информации с «термодинамической» энтропией; как атрибута материи или воплощения идеального. Обращение к теории информации не упрощает ситуацию, поскольку во всех ее дисциплинарных разделах саму информацию понимают предметно. Так, в математической теории связи информация отождествлена с количеством информации, а в динамической теории информации – с ценностью информации. Все направления теории информации объединяет то, что каждый раз, когда мы имеем дело с процессами выбора или отбора, мы имеем дело с информацией. Механизмы этих процессов в системах любой природы стали понятны благодаря достижениям фундаментальных наук в той области, которую в отечественных гуманитарных исследованиях принято называть синергетикой. С позиций синергетики, обозначенных (3 и 4) в вышеперечисленных основаниях, открылась методологическая возможность понимать феномен информации как процесс, состоящий из конкретных стадий, каждая из которых является предметом различных теорий информации, а итоги ключевых стадий понимать в качестве создания семиотических форм (1 и 2). Таким образом,

процессуальная трактовка феномена информации устанавливает связь синергетика-семиотика и предоставляет возможность привлечь характеристики информации если не для решения проблемы измерения в гуманитарных исследованиях, то для диагностики спектра аттракторов социокультурной динамики.

Подходы и методы семиотической диагностики

Суть подхода состоит в обнаруженной корреспонденции между фазами самоорганизации и стадиями информационного процесса [7], что открывает методологическую возможность измерения самоорганизации социокультурных систем на основе характеристик информации. В теории информации таких характеристик три (ценность, количество, эффективность), поэтому для осуществления процедур семиотической диагностики акцентирована корреспонденция между тремя фазами самоорганизации и тремя стадиями информационного процесса.

В теории информации есть различные аналитические виды для расчета характеристики «ценность информации». Ее расчет основан на понимании этой характеристики как изменения вероятности достижения цели. Это значит, что кривая $V = f(I)$, где V – ценность, а I – количество информации, для конкретной целевой ситуации будет иметь вид кривой с насыщением. Подобные кривые демонстрируют, что по ходу кривой будет изменяться приращение ценности, приходящейся на единичный интервал I , а с определенных значений I повышение значений V прекратится. Из этого следует, что первая производная функции $V = f(I)$ будет представлена кривой с максимумом, причем только с одним. Это позволяет ввести еще одну характеристику эффективности информации $\varepsilon = dV / dI$ [7, с. 26]. Необходимо подчеркнуть, что для разных множеств однотипных реципиентов одна и та же информация будет иметь различную ценность, а изменение ценности информации будет иметь различную эффективность. Это значит, что существует связь между тремя характеристиками информации, каждая из которых становится основной для описания разных стадий информационного процесса: генерации (V), трансляции (I) и реализации информации (ε):

1) преодоление системой хаотического состояния, движимое целью обрести устойчивое состояние, что сводится к выбору варианта «выхода из хаоса», подобного процессу генерации информации как случайному и запомненному выбору варианта;

2) распространение микросостояний «нового порядка» по всему объему системы, аналогичное процессу перехода, детерминированного исходным выбором, что приводит к изменению структурных уровней системы (усложнение при самоорганизации или упрощение при самодезорганизации), подобного формированию каналов трансляции и процессу передачи информации;

3) формирование спектра устойчивых состояний, достижение которых соответствует процессу отбора способов реализации информации.

Подчеркнем, что стадии (1) и (2) целенаправленны. Именно с позиции достижения некоей цели измеряют ценность выбора и эффективность отбора.

Подход основан на философии процесса А. Н. Уайтхеда, утратившей свою метафизичность в интерпретациях теории самоорганизации открытых сложных систем. Эти же интерпретации сделали понятной чередование двух видов процессов, установленных А. Н. Уайтхедом. Причиной одного из них есть цель, а причиной второго вида процесса есть начальные условия, сформированные по завершении первого вида процесса. Каждый вид процесса имеет свои изоморфные стадии. Процесс, движимый целевой причиной, есть совершение события, детерминирующее разворачивание процесса перехода к следующему событию, которое может стать управляемым при условии умения совершать «революции в символизме». Цель этой «революции» можно счесть метафорой «странного аттрактора», что в данной статье получит трактовку «семиотический аттрактор».

В философии процесса А. Н. Уайтхеда совершение события происходит при последовательной смене этапов: «the responsive phase», «the supplemental stage» и «the satisfaction» [4, p. 323], которые мы обозначим (i), (ii), (iii).

Релевантный метод исследования смены этапов найден в модификациях модели Лотки-Вальтерра [8]. Модель описывает процессы конкуренции, что соответствует конкуренции вариантов «выхода из хаоса» и осуществлению окончательного выбора одного из вариантов («the satisfaction»). В ходе осуществленных компьютерных экспериментов с параметрами модели был установлен ряд обстоятельств:

а) достичь «сатисфакции» способен любой из конкурирующих вариантов; б) отслеживание в реальном времени процессов конкуренции позволяет прогнозировать переходы (i)→(ii) и (ii)→(iii), но никогда переход от старта к финишу, то есть (ii) полностью соответствует

описанию А. Н. Уайтхеда в качестве «скалярной» [4, р. 323] и необходимой «дополнительной» стадии;

в) визуализация пространства решений дает изоморфную конфигурацию кластеров, в которые объединяются носители различных вариантов устойчивых состояний, и эта кластеризация происходит через одинаковые периоды времени, что получило интерпретацию в качестве структур-аттракторов;

г) носители различных вариантов группируются в кластеры практически мгновенно (i), а длительность формирования (iii) превосходит длительность предшествующего этапа (ii) на порядок.

Выясненные обстоятельства позволили дать семиотическую трактовку происходящего на этих стадиях: на первом этапе (i) фиксируется семантика или сам выбор варианта, на этапе (ii) – синтактика конкуренции вариантов, на этапе (iii) – прагматика варианта. Если учесть возрастающую длительность этапов и то обстоятельство, что весь процесс конкуренции движим целевой причиной, то семиотическую интерпретацию можно уточнить: на старте конкурируют семантические выражения цели, на финише – прагматические, и выигрыш в конкуренции одерживает тот вариант, у которого сиюминутно возникшая семантика кодирует близкую цель, а сатисфакцию получает вариант, кодирующий прагматику удаленной цели. Случайность переходов (i)→(ii) и (ii)→(iii) свидетельствует о том, что семантика выбора еще не означает выбор ценного варианта, то есть варианта, ведущего к достижению цели. Содержание же характеристики «эффективность информации» свидетельствует о том, что выбор ценного варианта еще не детерминирует эффективного отбора способа реализации избранного сценария. Иными словами, итог, ставший результатом процесса генерации информации, и итог, ставший результатом процесса отбора способов реализации информации, не являются одним и тем же «предметом». Весь информационный процесс есть путь создания и перевоплощения семиотических форм, в которых выражают себя результаты каждой его стадии. Этот путь вариативен, поскольку условие старта стадии зависит от финиша предшествующей стадии.

Данная траектория «роста символов», ведущая к «революциям в символизме», которые управляют жизнью людей, может быть оценена с позиций асимптотических целей социокультурной динамики. По этим целям можно задать оси для выражения координат, в которых происходит смена фазовых состояний социокультурных систем. Например, в самобытных эпохах этих систем понимание блага

воплощало себя как милосердная любовь, самосовершенствование человека, наслаждение и польза [9]. Эти воплощения диктовали свои способы правильной жизни людей, в которых находила реализация понимания блага. Альтруизм диктует подчинение личных стремлений интересам других людей в ситуации зависимости от того, что все понимают в качестве блага. Гедонизм же «разрешает» человеку активность и свободу, то есть видеть приоритет в достижении личных целей и выбирать свое понимание блага. Всё множество устремлений сводится к четырехмерному пространству: активность/свобода, активность/зависимость, пассивность/свобода, пассивность/зависимость [10]. Пространство является четырехмерным, поскольку «зависимость» не представляет собой область отрицательных значений «свободы». В этом пространстве координат можно обнаружить спектр «семиотических аттракторов», на которые укажет «вектор единичной окружности», если его начало поместить в пересечение осей, а «острие» определить по значениям характеристик информации, фиксирующих семиотические компоненты в кругах «революций в символизме», которые угадывает биоэтика, выставляя заслоны модификациям природы человека.

Семиотическая диагностика аттракторов конвергентных технологий

Распространено убеждение в том, что лучший способ реализации – тот, который представляет самую короткую траекторию достижения цели, то есть обладает максимальной эффективностью. Это убеждение справедливо для конструирования любых искусственных объектов. Технологии конструирования должны быть оптимальными, алгоритмы – лаконичными.

Слово «лаконичный» оставлено нам в наследство Спартой, всегда стремящейся к простым решениям. Насколько это является эффективным в преследовании асимптотических целей – иллюстрирует известный пример. В VIII - V веках до нашей эры все полисы Древней Греции столкнулись с проблемой аграрного кризиса, которую каждый полис решал самостоятельно. Первой ответ был найден Спартой: если чего-то не хватает, то это «чего-то» можно отнять. Отбор оптимальных способов достижения цели привел к формированию того, что много позже было названо милитаризмом. Второй вариант преодоления кризисной ситуации был предложен Коринфом: если чего-то не хватает, то это «чего-то» можно произвести в других условиях, что выражает колониальную политику. У Афин на поиск ответа ушло

много времени, но еще больше времени потребовало создание способов реализации этого ответа: если чего-то не хватает, то это «чего-то» можно купить. Разработка способа, предложенного Афинами, стала «Школой Эллады», привела к резкому взлету всех составляющих благополучия полиса. Этот способ стал зародышем демократии и рыночной экономики. Справедливости ради надо вспомнить, что Афины, искрящиеся всеми проявлениями богатства, были позже разбиты унылой и бедной Спартой. Итак, всегда ли получают «сатисфакцию» лаконичные способы? Как отличить «поражение от победы»?

Ответ на первый вопрос дает модель «семиотического аттрактора», демонстрирующая, что стадии «сатисфакция» может достичь любая инновация. Этот результат драматичен, поскольку асимптотической целью может завладеть инновация, которая обладает лишь конъюнктурной ценностью. Необратимость процесса компенсирует некое «торможение», происходящее на промежуточной стадии, поэтому движение к «сатисфакции» можно корректировать. Но платой за эту корректировку будет неизбежный возврат к исходной проблемной ситуации, уже отягощенной негативными последствиями, и потеря темпа инновационной динамики, что противоречит предназначению когнитивного менеджмента. Этот вывод обладает острой актуальностью для менеджмента в сфере NBIC-технологий. Поскольку NBIC-технологии делают реальным осуществление произвольных манипуляций над силами природы, телесностью человека и его осознанием себя, что деформируют духовные иммунитеты, полагавшие пределы созданию и применению технологий в конструировании человека человеком. Энтузиазму NBIC-технологий оппонирует биоэтика [1-2, 10], что и выполняет функцию «торможения». Вмешательство биоэтики диагностирует этап, на котором в настоящее время находится социокультурная динамика, вызванная конвергентными технологиями. Этай стадии («the supplemental stage»), как подчеркивал А. Н. Уайтхед, свойственна «скалярная» форма, преобладающая над изначально «векторной» формой, и, как следует из результатов проведенных компьютерных экспериментов, свойственно доминирование синтактики в символизации целей. Эту диагностику подтверждает доминирование визуальных форм, в которых выражают свои устремления новые технологии. Аббревиатура в названии технологий фиксирует первенство нанотехнологий, инициированных изменением фундаментальных представлений о возможностях управления видимым для человеческого глаза электромагнитным

излучением, но первенство принадлежит информационным технологиям, инициированным открытым кибернетикой сходством в управлении биологическими и искусственными системами. Данные системы «предпочитают» лаконичные способы управления, что является необходимым условием для разработки любых технологий – создавать максимально простой способ достижения запланированного результата и редупликации этого результата в любых обстоятельствах. Продолжающаяся конвергенция новых технологий увеличивает их шансы достичь «сатисфакции», то есть одержать победу искусственного совершенства над естественными несовершенствами человека.

Ответ на второй вопрос дает семиотическая диагностика, проводимая с позиции биоэтики. Если технологии направлены на воссоздание утраченной нормы, то это действие имеет полное моральное обоснование. Но если технологии направлены на превосходство нормы, то в этом усматривают манипуляцию, имеющую моральную неопределенность, поскольку есть опасность стандартизировать уникальную индивидуальность. Проведение границы между этими действиями затруднено скольжением пределов, в которых понимают саму норму. Пока область устремлений конвергентных технологий напоминает Вавилонскую башню. Притча изложена в Книге Бытия в 9 предложениях. В этой краткости отведено место техническому решению проекта: «и стали у них кирпичи вместо камней, а земляная смола вместо извести» [Быт. 11:3. 3, с. 17]. Метафору о Вавилонской башне обычно понимают как наказание людей лингвистической разобщенностью за гордый замысел. Но либо срок наказания истек, либо не оценено дарованное избавление от занятия, напрасного и уравнивающего всех, только мудрость слов «не отстанут они от того, что задумали сделать» [Быт. 11:6. 3, с. 17] подтверждают предпринимаемые грандиозные проекты и единение разработчиков проектов с потребителями их продуктов. Приведенная метафора призвана зародить сомнение в действительном социокультурном усложнении, вызываемом инновациями. Ведь пока в их итогах – родство профанов, поскольку каждый разработчик конкретной технологии оказывается в беспомощности непонимания всех остальных технологий; упрощение языка коммуникаций, предоставляемое «паутиной» и основанное на двоичном коде «цифровой культуры»; и устремленность к коллективным мечтам, продвигаемых социальными технологиями, участие которых иногда фиксируют в продолжении аббревиатуры

конвергентных технологий. Итак, в перспективе одна роль, один язык, одна мечта о жизни на «башне». Отсутствие точных критериев диагностики направлений динамики делает умозрительной гуманитарную экспертизу новых технологий, управляющих нашей жизнью и впечатлениями о ней. Иллюзии и отсутствие средств декодирования властной символики уже приводят к тому, что в биомедицинских технологиях происходит утрата индивидуальности, в образовательных технологиях – личности, а в философской рефлексии социокультурных трансформаций – субъекта. Мы будем бесконечно строить и разрушать подобию Вавилонской башни, совершенствуя технологии их возведения и уничтожения. Мы будем принимать растущую множественность деталей за сложность и разнообразие, образы коллективной мечты за символы асимптотических целей, деформации социальной памяти за прогресс интеллектуальных традиций, а уравнивающие нас технологии, специфика которых в упрощенном способе воспроизводства результата, за способы обретения родства и могущества.

ЛИТЕРАТУРА

1. *Illes J., Bird S.* Neuroethics: a modern context for ethics in neuroscience // *Trends Neurosci.* 2006. Sept., 29 (9), p. 511-517
2. *Юдин Б. Г.* Человек как объект технологического воздействия // *Человек.* 2011. № 3. С. 5-20
3. *Пирс Ч. С.* Что такое знак? // *Вестник Томского государственного университета. Философия. Социология, Политология.* 2009. № 3. С. 88-95.
4. *Whithehead A. N.* Process and Reality. – N. Y: The Macmillan Company, 1929.
5. *Пригожин И. Р., Стенгерс И.* Время, хаос, квант. – М.: Прогресс, 1984.
6. *Князева Е. Н., Курдюмов С. П.* Синергетика как новое мировидение: диалог с И. Р. Пригожиным // *Вопросы философии.* 1992. « 12. С. 3-20.
7. *Мелик-Гайказян И. В.* Информационные процессы и реальность – М.: Наука. Физматлит, 1998.
8. *Evdokimov K. V., Kuznetsov V. G., Melik-Gaykazyan I. V., Melik-Gaykazyan M. V.* Semiotic Interpretation of Lotka-Volterra Model and its Usage in Knowledge Management. // *SHS Web of Conferences*, 2016. DOI: 10.1051/shsconf/20162801033
9. *Апресян Р. Г.* Ценностные парадигмы воспитания // *Вестник Томского государственного педагогического университета.* 2008. № 1. С. 89-94.
10. *Melik-Gaykazyan I. V., Melik-Gaykazyan M. V., Mescheryakova, T. V., Sokolova N. S.* Model of Bioethics as «Semiotic Attractors» for Diagnosing Innovative Strategies of Training Specialists for NBICS-Technologies Niche. // *SHS Web of Conferences*, 2016. DOI: 10.1051/shsconf/20162801069

«ВИТГЕНШТЕЙНИАНСКИЙ» ИНТЕНЦИОНАЛЬНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ WITTGENSTEINIAN INTENTIONAL ARTIFICIAL INTELLIGENCE

瓦锈可副 ««维特根斯坦»意图人工智能»

**Васюков Владимир
Леонидович**

*доктор философских наук, заведующий кафедрой
истории и философии науки Института
философии РАН, г. Москва*

Vasyukov Vladimir L.

*Doctor of Philosophy, Head of the Department
of History and Philosophy of Science, Institute of
Philosophy RAS, Moscow*



Аннотация. В статье описывается, как мы можем использовать идеи Л. Витгенштейна в системах искусственного интеллекта. Показано, как на этом пути можно преодолеть тупики хайдеггерианского искусственного интеллекта, основанного на идеях М. Хайдеггера. Модифицируя понятие интенциональности, данное Р. Дж. Серлем, описывается гипотетическая модель витгенштейнианского интенционального искусственного интеллекта, основанная на комбинациях логических и дедуктивных систем.

Ключевые слова: искусственный интеллект, интенциональность, комбинации логических систем, интенциональная установка, психофизическая проблема.

Annotation. The article describes how we can use L. Wittgenstein's ideas in artificial intelligence systems. It is shown how on this way it is possible to overcome the dead ends of the Heideggerian artificial intelligence, based on the ideas of M. Heidegger. Modifying the concept of intentionality given by R.J. Searle, describes a hypothetical model of Wittgensteinian intentional artificial intelligence, based on combinations of logical and deductive systems.

Keywords: artificial intelligence, intentionality, combinations of logical systems, intentional installation, psychophysical problem.

Хьюберт Дрейфус в своей работе «Почему хайдеггерианский ИИ потерпел неудачу и как его улучшение может потребовать еще большей хайдеггеризации»¹ описывает историю разработок систем искусственного интеллекта (ИИ), которую кратко можно представить в виде последовательности следующих этапов.

Когнитивная симуляция

Ньюэлл и Саймон провозглашают, что цифровые компьютеры, как и человеческое сознание, можно понимать как систему физических символов, используя последовательности битов или потоки нейронных импульсов как символы, отображающие внешний мир. Интеллекту требуется просто делать соответствующие выводы из этих внутренних символизаций. По сути дела, разработчики ИИ этого этапа исходили из гоббсовского утверждения, что мыслить означает вычислять, декартовских ментальных представлений, лейбницеvской идеи «универсальной символизации», кантовского полагания, что понятия суть правила, фрегевской формализации этих правил и расселовского постулирования логических атомов в качестве строительных блоков реальности. Без всего этого трудно было бы основывать исследовательскую программу ИИ на рационалистической философии.

Упадок программы символического ИИ

Однако разработки ИИ, наряду с другими проблемами, столкнулись с проблемой репрезентации смысла и релевантности – проблемой, которую Хайдеггер подметил у Декарта в его имплицитном понимании мира как множества бессмысленных фактов, к которым сознание привязывает то, что Декарт называл значимостью (ценностями), а Джон Сёрль в наше время называет функциями. Хайдеггер указывает, что занимаясь лишь значимыми (ценностными) характеристиками, мы не в состоянии понять такой аспект бытия, как «подручность», когда, например, «подручность» является способом бытия молотка и поддающаяся интерпретации смысловая организация повседневного

Статья подготовлена в результате проведения исследования (№15-05-0005) в рамках Программы «Научный фонд Национального исследовательского университета 'Высшая школа экономики' (НИУ ВШЭ)» в 2016 г. и в рамках государственной поддержки ведущих университетов Российской Федерации «5-100».

¹ Dreyfus H. L. Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian // Philosophical Psychology 20 (2), 2007. P.:247 – 268.

мира включает в себя обработку молотком. Здесь возникает проблема, как действовать компьютеру, не имея представления о релевантности его знаний о состоянии мира. Марвин Минский предположил, что для избавления от подобной проблемы компьютер должен использовать «фреймы» – описания типичных ситуаций, когда создается список и используются те и только те факты, которые действительно релевантны. Но как компьютеру выбрать из миллиона фреймов, хранящихся в его памяти, релевантные для мира фреймы? Чтобы решить эту проблему, в 60-х и начале 70-х годов было решено ограничить компьютерные программы лишь микромирами – искусственными ситуациями с малым количеством свойств, из которых легко выбрать релевантные. Предполагалось, что по мере использования все более реалистичных фреймов удастся получить миры, отвечающее нашему здравому смыслу. Но этого не произошло. Не помог даже разработанный Терри Виноградом «язык представления знаний». Символический ИИ этого периода получил имя Старого Доброго ИИ (СДИИ).

Хайдеггерианский ИИ. Этап 1: элиминация репрезентаций путем построения поведенческих роботов

Первым дезертиром из программы СДИИ стал Виноград, который в свои лекции по компьютерным наукам стал включать философию Хайдеггера. Имплицитно включающим в себя хайдеггеровскую критику декартовских внутренних репрезентаций стал также бихевиористский подход Родни Брукса. Брукс, основываясь на идее «лучшей моделью мира является он сам», разработал мобильных роботов, которые использовали сам мир в качестве своей репрезентации, постоянно адресуясь к своим сенсорам, а не к внутренней модели мира. Они реагировали только на стабильные свойства окружающей среды, не обучались и были похожи в этом отношении на муравьев. Брукс достаточно критично оценивал возможности своих роботов, однако это не помешало ему вместе с Дэниелом Деннетом попытаться перепрыгнуть этапы совершенствования «муравьев» (например, построить искусственных пауков), перейдя к созданию гуманоидного робота, который мог бы говорить, манипулировать объектами с помощью визуальной координации и демонстрировал бы самосохранение, самоуправляемость и самоанализ. Этот проект был рассчитан на длительное время,

но оказался короткоживущим. Его главной ошибкой было то, что Иешуа Бар-Хиллел на заре разработок ИИ назвал ошибкой первого шага – считать, что первая успешная попытка обезьяны взобраться на дерево была первым шагом к полету на Луну.

Хайдеггерианский ИИ. Этап 2: программируя подручность

Новый этап разработок ИИ начался после того, как в своем докладе в Массачусетском технологическом институте Дрейфус познакомил разработчиков с хайдеггеровским нерепрезентационистским объяснением погруженности Dasein (человеческого бытия) в мир и хайдеггеровским выделением двух типов бытия: подручности инструментов, когда мы используем их, и всего-лишь-наличием, когда мы их рассматриваем. Это непосредственно повлияло на последующие разработки ИИ. Программа Фила Эгре (Phil Agre) и Дэвида Чапмена с виртуальным агентом, играющим в компьютерную игру «Пенго», когда агент и пингвин забрасывают друг друга кусками льда, основывалась на «интеракционистском» подходе и была гораздо более хайдеггерианской, чем программы Брукса. Эгре использовал «дейктические репрезентации», основывающиеся на прямой аналогии с хайдеггеровским анализом повседневной интенциональности. Эти репрезентации передают не конкретные объекты мира, но скорее роль, которую они могут играть в некоторой длящейся во времени конфигурации взаимодействия между агентом и его окружением.

Псевдохайдеггерианский ИИ: встроенное, овеществленное, расширенное сознание

Следующим этапом и одновременно шагом в сторону стала разработанная Уолтером Фриманом когнитивная модель мозга как динамической физической системы, показывающая, что мозг делает для того, чтобы обеспечить себя материальным субстратом хайдеггеровского и мерло-понтианского объяснения обыденного восприятия и деятельности. В модели Фримана состояния являются репрезентациями состояний мозга, а не свойствами вещей в повседневном мире, она является не интенциональной системой, но лишь ее описанием. К сожалению, Фриману не удалось достроить до конца свою когнитивистскую парадигму, хотя подход мог бы стать шагом в построении подлинно экзистенциального ИИ. Тем не менее Дрейфус считает модель Фримана основанием хайдеггерианского ИИ.

Дрейфус заканчивает историю хайдеггерианского ИИ рассмотрением вопроса о том, как хайдеггерианский ИИ может устранить,

а не избежать или решить проблему фреймов. Здесь он присоединяется к мнению Майкла Уилера, что критерием жизнеспособности любой предложенной программы ИИ является ее способность разрешения проблемы фреймов. В описанной выше истории по крайней мере два хайдеггерианских подхода к ИИ скорее избегали, чем решали эту проблему. Подход Брукса, когда реакция вызывается непосредственно окружением, избегает ее, вообще не рассматривая значение и обучение, а подход Эгре, включающий лишь ограниченное количество возможных релевантных реакций, не учитывает проблему меняющейся релевантности. Однако и собственное предложение Уилера включения гибких, ориентированных на действия репрезентаций, как и любой репрезентационный подход, сталкивается с той же проблемой фреймов.

Напомним, в чем заключается проблема фреймов. Фреймы представляют собой описания типичных ситуаций, когда создается список и используются те и только те факты, которые действительно релевантны. Проблема состоит в выборе компьютером релевантных фактов из миллиона фреймов, хранящихся в его памяти. Чтобы решить эту проблему, компьютерные программы ограничивали лишь искусственными ситуациями с малым количеством свойств, из которых легко выбрать релевантные. Миры, отвечающие нашему здравому смыслу, получаются по мере использования все более реалистичных фреймов.

Таким образом, проблема фреймов в первую очередь связана с резервами памяти компьютера, не позволяющими ему оперировать всеми фактами и фреймами для выбора релевантных фактов. Ведь даже имея хорошо сформулированные условия релевантности, трудно, в принципе, остановиться в своем выборе, обращаясь все к новым и новым ситуациям. Соблазн «перфекционизма» здесь достаточно силен. Вдобавок глобальный мир (суперфрейм), отвечающий нашему здравому смыслу, будет образован бесконечным множеством фактов.

Однако вспомним, что пишет Людвиг Витгенштейн в «Логико-философском трактате»: «1.1. Мир есть совокупность фактов, а не предметов» и «1.11. Мир определен фактами и тем, что это все факты»¹. Совладать с этим бесконечным миром фактов можно, если использовать логику: «5.61. Логика наполняет мир; границы мира являются

¹ Витгенштейн Л. Логико-философский трактат. М.: Канон+, 2008. С. 36.

также ее границами»¹. И далее: «6.13. Логика не теория, а отражение мира. Логика трансцендентальна»² и «6.24. Логические предложения описывают строительные леса мира, или, скорее, изображают их»³. Отсюда следует, что использование логических конструкций обеспечивает не только наличие глобального суперфрейма (что дается всей логикой), но вдобавок позволяет «изобразить» конкретные структуры мира, смоделировать восприятие и деятельность.

На первый взгляд, кажется, что логический подход трудно согласовать с контекстуальной зависимостью, присущей человеческой памяти. Фриман, описывая свою когнитивную модель мозга, которую, по мнению Дрейфуса, можно считать основанием хайдеггерианского ИИ, писал, что «контекстуальная зависимость является существенным свойством интеллектуальной системы памяти, в которой каждое новое восприятие должно изменить все содержание памяти малой порцией восприятия, чтобы новое поступление было встроено и полностью использовано в существующей системе восприятий. Это свойство контрастирует со структурой памяти в компьютерах, в которых каждый элемент данных определяется адресом или ветвью дерева поиска. Там каждая единица данных хранится в ячейке и новые элементы не меняют старых. Наши данные показывают, что в мозге память не имеет никаких ограничений или ячеек... каждое новое состояние перехода ... инициирует конструкцию локального шаблона, который внедряется и модифицирует всю интенциональную структуру»⁴. Но если вспомнить системы натурального вывода⁵, в которых вывод, например, имплицативных формул приводит к блокировке подвыводов и невозможности в дальнейшем использовать заблокированные формулы, то это как раз отвечает той части фримановского описания, когда «каждое новое состояние перехода ... инициирует конструкцию локального шаблона, который вторгается и модифицирует всю интенциональную структуру».

К тому же, даже мерло-понтианское объяснение обыденного восприятия и деятельности, о котором говорил Дрейфус, также можно

¹ Там же, с. 174.

² Там же, с. 196.

³ Там же.

⁴ Freeman W. Societies of Brains: A study in the neuroscience of love and hate // Hillsdale, N.J. (ed.). The Spinoza Lectures, Amsterdam: Lawrence Erlbaum Associates, 1995. P.99..

⁵ См., напр., Смирнов В. А. Логический анализ научного знания М.: УРСС, 2001, с. 27.

«изобразить» логически. Морис Мерло-Понти в «Феноменологии восприятия» пишет: «жизнь сознания — познающая жизнь, жизнь желания, или жизнь перцептивная — скрепляется «интенциональной дугой», которая проецирует вокруг нас наше прошлое, будущее, наше житейское окружение, нашу физическую, идеологическую и моральную ситуации или, точнее, делает так, что мы оказываемся вовлечены во все эти отношения. Эта интенциональная дуга и создает единство чувств, единство чувств и мышления, единство чувствительности и двигательной функции»¹. Ключевым моментом здесь является понятие «интенциональности», которое, если верить Витгенштейну, тоже можно рассмотреть или изобразить, используя логические «строительные леса».

Понятие «интенциональности» для философов не ново, достаточно вспомнить, что про интенциональности говорил еще Ф. Brentano, который в своей работе «Психология с эмпирической точки зрения» писал, что «всякий психический феномен характеризуется ... направленностью на объект ... В представлении нечто представляется, в суждении нечто утверждается или отрицается, в любви — любит-ся, в ненависти — ненавидится, в желании желается и т. д.». Каждый акт сознания имеет свой предмет, направлен на предмет. Отношение между актом и его субъектом есть интенциональное отношение.

В дальнейшем в философии сознания понятие интенциональности претерпело значительные изменения. Дж. Сёрль в этой связи пишет: «Интенциональность есть то свойство многих ментальных состояний и событий, посредством которого они направлены на объекты и положения дел внешнего мира. Если, например, я верю, то это должна быть вера в то, что нечто имеет место; если я боюсь, то я боюсь чего-то; если я хочу что-то сделать или хочу, чтобы что-то произошло; если у меня имеется некоторое намерение, то это намерение что-то совершить»². Правда, он тут же добавляет: «во многих отношениях данный термин является ошибочным и вносит путаницу, поэтому я с самого начала хочу ... отмежеваться от некоторых особенностей указанной традиции»³.

Путаница возникает потому, что далеко не все ментальные состояния и события являются интенциональными, например, существуют

¹ Мерло-Понти М. Феноменология восприятия. СПб.: «Ювента» «Наука» 1999. С. 182.

² Сёрль Дж. Р. Природа интенциональных состояний // Философия. Логика. Язык. М., 1987. С. 96.

³ Там же.

формы восторга и уныния, которые не имеют конкретного повода, но они интенциональны, когда чем-то вызваны. Кроме того, интенциональность не тождественна осознанности (обычно это принимается как очевидное: если вы направляете ваше сознание на что-то, то вы всегда делаете это осознанно, само ваше сознание есть сознание чего-то). Если я боюсь тигра, то переживание страха и страх тождественны, но сама боязнь тигра не тождественна тигру. Наконец, намеренность (интенция) сама по себе также является всего лишь одной из форм интенциональности и не имеет особого статуса. Она ничем не отличается от веры, надежды, страха, желания и т.п.

Не следует путать интенциональные состояния и речевые акты: ментальные состояния являются интенциональными состояниями, а речевые акты – интенциональными действиями. Конечно, в речевом акте присутствует интенциональное состояние, но вместе с тем имеется и интенция как намерение, с которым что-то произносится. Эта интенция и наделяет интенциональностью физические феномены.

В этой связи можно легко различать логические свойства интенциональных состояний и их онтологический статус. Формы реализации интенционального состояния столь же безразличны для его логических свойств, как формы реализации речевого акта безразличны для логических свойств этого акта. Интенциональный объект не имеет какого-то особого онтологического статуса, назвать что-то интенциональным объектом означает просто сказать, что это тот объект, к которому относится некоторое интенциональное состояние.

Интенциональные состояния по Серлю состоят из содержания и психологических модальностей, и этими содержаниями часто являются суждения. Он пишет: «Интенциональные состояния обычно обладают структурой, аналогичной структуре речи. К примеру, я могу приказывать вам покинуть комнату, спросить, не покинете ли вы комнату, и предположить, что вы уйдете из нее; так же я могу надеяться, что уйдете из комнаты, бояться, что вы уйдете из комнаты, или желать, чтобы вы покинули комнату. В каждом случае существует пропозициональное содержание, заключающееся в том, что вы уйдете из комнаты. Это содержание проявляется в тех или иных лингвистических или психологических модальностях. В языковом отношении оно может, например, принимать формы вопроса, предсказания, обещания или приказа. А в случае мышления оно может, например, принимать форму убеждений, страхов и желаний»¹.

¹ Сёрль Дж. Р. Рациональность в действии. М., 2004, с. 52-53.

Сёрль описывает общую структуру интенциональности как $S(p)$,

где S обозначает тип психологического состояния, а p – пропозициональное содержание этого состояния. Их следует различать, поскольку одно и то же пропозициональное содержание может появляться в разных психологических модальностях. Интенциональные состояния с пропозициональным содержанием могут либо соответствовать реальности, либо не соответствовать ей. То, как они должны соответствовать реальности, как раз и определяется психологической модальностью.

С точки зрения логики серлевская формула интенциональности есть не что иное, как второпорядковое предсказание, когда психологическое состояние является некоторым свойством пропозиционального содержания. Второпорядковое описание призвано имитировать корреляцию ментальных и физических состояний мозга, если вспомнить проблему психофизического дуализма. Оно призвано преодолеть ту лежащую в основании когнитивистики ошибку, на которую указывал Сёрль, заключающуюся в том, что мозг рассматривался когнитивистами с самого начала как компьютер, а сознание – как компьютерная программа.

Как пишет Сёрль, «сознание не может представлять собой программу или программы, потому что синтаксические операции не являются сами по себе достаточными для образования или не могут быть гарантией присутствия семантического содержания фактических ментальных процессов. Сознание, с другой стороны, включает в себя не только символические или синтаксические компоненты, оно содержит фактические ментальные состояния с семантическим содержанием в виде мыслей, чувств и т.д., и они обусловлены специфическими нейробиологическими процессами в мозгу. Сознание не может быть программой, потому что синтаксические процессы осуществляемой программы сами по себе не обладают никаким семантическим содержанием»¹.

Однако в случае ИИ мы как раз и имеем дело с сознанием-программой. И в этом случае серлевское определение интенциональности должно быть модифицировано. Оно должно учитывать «параллельное», но не независимое прочтение программы, функционирование синтаксических процессов в компьютере, имитирующее

¹ Searle J. R. *Philosophy in a New Century*. N.Y.: Cambridge University Center, 2008. P. 15.

психофизический дуализм. Определение интенциональности с помощью предсказания было следствием стремления теоретиков интенциональности рассматривать более общую концепцию сопряжения, корреляцию психического и физического. Однако в логике существуют и иные, более абстрактные способы связывания между собой логических свойств и аспектов.

Одним из подобных способов является конструкция произведения логических систем, которую можно использовать с данной целью¹. Логическую систему, полученную с помощью произведения двух других систем, можно охарактеризовать как такую логику, в которой «формулы» представляют собой пары формул, а следствия получаются путем комбинирования соответствующих следствий из двух исходных систем. Формально это означает, что формулы выглядят как пары $\langle A_1, A_2 \rangle$, в то время как $\langle B_1, B_2 \rangle$ следует из $\langle A_1, A_2 \rangle$ тогда, когда B_1 является следствием A_1 в рамках своей системы, а B_2 является следствием A_2 в рамках своей.

Следование $A_1 \vdash B_1$ в нашем случае было бы призвано моделировать каузальную связь физических состояний, в то время как следование $A_2 \vdash B_2$ должно пониматься как ментальное следование, связь ментальных состояний, отличную от каузальной связи, т.е. психофизический дуализм моделируется связыванием этих двух разновидностей следования в одно, точнее в пару следований $\langle \vdash_1, \vdash_2 \rangle = \langle \vdash_1 \otimes \vdash_2 \rangle$. Подобная техника позволяет для пары подобных «логик» найти некоторую «объединенную» логическую систему, которая будет обладать всеми свойствами, детерминированными рассматриваемой конструкцией пар формул и пары отношения следования.

В рамках подобного подхода мы гипотетически можем описать даже некую разновидность «интенциональной интеллектуальной интуиции», рассматривая последовательность выводов $\langle A_1, B_1 \rangle \vdash_1 \otimes \langle A_1, B_2 \rangle$, $\langle A_1, B_2 \rangle \vdash_1 \otimes \langle A_1, B_3 \rangle$, ... как описание ситуации «мгновенного суждения», когда психологическое состояние фиксировано, т.е. $A_1 \vdash_1 A_1 \vdash_1 A_1$, ... , а пропозициональное содержание изменяется таким образом, что у нас возникает последовательность выводов $B_1 \vdash_2 B_2$, $B_2 \vdash_2 B_3$, ...

Наоборот, последовательность выводов $\langle A_1, B_1 \rangle \vdash_1 \otimes \langle A_2, B_1 \rangle$, $\langle A_2, B_1 \rangle \vdash_1 \otimes \langle A_3, B_1 \rangle$, ... дает нам описание «обратной интенциональности», когда фиксированное пропозициональное содержание

¹ Васюков В. Л. Проблема контекста интерпретации в универсальной логике // Логические исследования. Вып. 14. М., 2007. С. 105-130.

вызывает смену психологических состояний в виде последовательности выводов $A_1 \vdash_2 A_2, A_2 \vdash_2 A_3$ и т.д. Это можно было бы понимать как ответ на вопрос, каким образом наша мыслительная деятельность может влиять на каузальную связь физических состояний, т.е. как описание механизма «ментального» управления телесными свойствами, обратную связь психики с телом.

Этот же механизм можно описать и иначе, используя конструкцию так называемых экспоненциалов логических систем. Система-экспоненциал двух логических систем возникает тогда, когда существуют два взаимно обратных перевода h, g из одной логики в другую и результаты всех переводов из одной системы в другую и обратно навязывают свою связь по следованию формулам исходной системы. В этом случае формулы A и B будут следовать друг из друга, когда их результаты переводов $hg[A]$ и $hg[B]$ будут связаны между собой отношением следования первой системы.

При наличии подобных переводов последовательность выводов $\langle A_1, B_1 \rangle \vdash_{1 \otimes 2} \langle A_2, B_2 \rangle, \langle A_2, B_2 \rangle \vdash_{1 \otimes 2} \langle A_3, B_3 \rangle \dots$ превращаются, например, в последовательность выводов $\langle A_1, B_1 \rangle \vdash_{(1 \Rightarrow 2) \otimes 2} \langle A_2, B_2 \rangle, \langle A_2, B_2 \rangle \vdash_{1 \otimes 2} \langle A_3, B_3 \rangle \dots$, когда $hg[A_1] \vdash_1 hg[A_2], hg[A_2] \vdash_1 A_3, \dots$ и $B_1 \vdash_2 B_2, B_2 \vdash_2 B_3$, где $\vdash_{(1 \Rightarrow 2)}$ означает вывод в экспоненциале наших двух логических систем. В этом случае мы получаем описание ситуации, когда ментальное следование влияет на каузальное следование, поскольку здесь каузальная выводимость определяется именно взаимнообратным сопоставлением каузальных и ментальных свойств с помощью композиции переводов hg .

Произвольность и многочисленность связей в данном механизме «обратной интенциональности» можно в какой-то степени преодолеть с помощью конструкции коэкспоненциала двух логических систем. В этом случае системе-коэкспоненциалу достаточно наличия одного перевода из одной логики в другую. Формулы A и B будут следовать друг из друга в коэкспоненциале, когда их результаты любого перевода $h[A]$ и $h[B]$ будут связаны между собой отношением следования второй системы.

При наличии подобных переводов последовательность выводов $\langle A_1, B_1 \rangle \vdash_{1 \otimes 2} \langle A_2, B_2 \rangle, \langle A_2, B_2 \rangle \vdash_{1 \otimes 2} \langle A_3, B_3 \rangle \dots$ превращаются, например, в последовательность выводов $\langle A_1, B_1 \rangle \vdash_{(1 \Leftarrow 2) \otimes 2} \langle A_2, B_2 \rangle, \langle A_2, B_2 \rangle \vdash_{1 \otimes 2} \langle A_3, B_3 \rangle \dots$, когда $h[A_1] \vdash_2 h[A_2], [A_2] \vdash_1 A_3, \dots$ и $B_1 \vdash_2 B_2, B_2 \vdash_2 B_3$, где $\vdash_{(1 \Leftarrow 2)}$ означает вывод в коэкспоненциале наших двух логических систем. В этом случае мы также получаем

описание ситуации, когда ментальное следование влияет на каузальное следование, поскольку здесь каузальная выводимость определяется сопоставлением каузальных и ментальных свойств с помощью перевода h , однако здесь сопоставление будет жестким в том смысле, что при любой связи (переводе) ментальное следование гарантировано. То есть, в этом случае речь идет о фиксации каузального следования любым ментальным выводом.

Программирование подобных разновидностей последовательностей выводов обещает привести к логическому описанию работы «витгенштейнианского интенционального ИИ» (витгенштейнианского ИИИ), когда присутствует два уровня сознания – каузальный (физический) и ментальный, представленные двумя логическими системами. В этой модели, по сути дела, происходит модификация серлевского представления интенциональности. Формула Серля переписывается в виде $\langle S, p \rangle$, точнее в виде совокупности пар $\langle S_i, p_i \rangle$, где S_i – психологические состояния, а p_i – соответствующие им пропозициональные содержания. Если же мы хотим в какой-то степени учитывать и подход нейробиологической когнитивистики, то следовало бы рассматривать не пары, а тройки $\langle N_i, S_i, p_i \rangle$, где третий элемент N_i представляет собой физические состояния нейрона. Аскриптивность (приписываемость) интенциональности понимается в этом случае как приписывание к состоянию нейрона пары $\langle S_i, p_i \rangle$.

Как следствие, в этом случае требуется работа уже не с двумя, но с тремя системами, и не с парами, но тройками формул и следований. Первые два следования – это каузальное и ментальное следования, а третье – следование в системе пропозициональных содержаний, и оно теперь будет зависеть от принимаемой логики, описывающей связь этих состояний. То, что это будет действительно некая система, можно утверждать на основании серлевского «тезиса Фона», когда интенциональные состояния не функционируют автономно и каждое интенциональное состояние нуждается в Сети других интенциональных состояний. Эта Сеть, в свою очередь, функционирует только относительно совокупности фоновых способностей, не связанных прямо с интенциональностью, что можно истолковывать как раз в смысле использования троек $\langle N_i, S_i, p_i \rangle$ с целью получения полного описания.

Само понятие интенциональности в витгенштейнианском ИИИ можно связать с позицией Д. Деннета, считавшего, что об интенциональности следует говорить не только в применении к человеческому сознанию. Нужно использовать интенциональную установку

всегда, когда требуется лучше и эффективнее объяснять и предсказывать поведение систем, допуская наличие в них цели и способности достигать эту цель. Деннетовская интенциональная установка применима только к интенциональным системам, чье поведение можно объяснить и предсказать, основываясь на приписывании системе неких «верований» и «желаний».

Используя подобную интенциональность в системах искусственного интеллекта, мы тем самым избегаем аналогий со сверхсложной проблемой природы интенциональности в реальном сознании, ее происхождении и функционировании. Деннету для объяснения сознания, интенциональности, способности к языку и производства значений, являющимися результатами эволюции, приходится говорить о том, что наша интенциональность является творением Главного Дизайнера – Матери-природы. Именно ей присуща первичная прото-интенциональность, наша же интенциональность вторична. Но нам этого делать не нужно, можно говорить о том, что это мы, заменяя собой деннетовскую Мать-природу, заложили «интенциональный» механизм в компьютер с целью моделирования работы сознания, пытаясь наделить искусственный интеллект «сознанием», схожим с человеческим. Задача теперь заключается только в механизме приписывания системе первичных и производных интенциональностей.

В нашем случае логику пар <элемент логики нейрона, элемент логики ментальных состояний> можно, по-видимому, рассматривать как некоторый вариант первичной протоинтенциональности. При этом мы явно неоправданно расширяем понятие интенциональности вообще, но оправдать это можно следующим образом. На аппаратном уровне своего функционирования компьютер оперирует с ячейками и сигналами, физический механизм функционирования которых сам по себе не имеет ничего общего с вычислениями (функций и т.д.). Считывая результаты, мы точно так же не интересуемся его физическими процессами и их направленностью. Даже работая с языками программирования низкого уровня, позволяющего непосредственно работать с ячейками памяти, мы используем процедуры работы с ячейками памяти для своих задач, а не для обеспечения физического режима работы деталей и частей компьютера (хотя, конечно, перегрузка памяти заставляет порой программиста изменить процедуру расчета).

Более того, мы и в этом случае говорим, что компьютеру «тяжело работать в таком режиме», наделяя компьютер какими-то свойствами

сознания. Логика наших вычислений диктуется законами математики, и поэтому мы приписываем компьютеру некоторую «математическую» интенциональность: наша интенциональная установка требуется здесь для того, чтобы лучше и эффективнее объяснять и предсказать поведение вычислительной системы, допуская наличие в них цели – результата вычисления – и способности достигать эту цель.

Аскриптивность интенциональности приводит к дальнейшему развитию этой протоинтенциональности – к вторичной интенциональности как возможности рассмотрения троек $\langle N_i, S_i, p_i \rangle$ в качестве средства описания и объяснения интенциональных аспектов работы компьютера, превращая его в систему искусственного интеллекта. В этом случае третий компонент берет на себя роль соответствующего паре $\langle N_i, S_i \rangle$ пропозиционального содержания. Вопрос лишь в том, что мы подразумеваем в качестве этого пропозиционального содержания.

Если дело в том, что мы «очеловечиваем» компьютер, наделяя его какой-то человеческой интенциональностью, то это тот же самый упомянутый выше случай, когда мы говорим о компьютере, что он «поразительно быстро, без сучка и задоринки выполнил сегодня программу вычислений», или, наоборот, в случае заикливания программы говорим, что «машина просто отказывается выполнять бредовые программы» и т.д.

Выход можно найти, прибегнув к рецепту А.Л. Блинова¹: воспользуемся притворными (игровыми) полаганиями насчет интенциональности пары $\langle N_i, S_i \rangle$, т.е. теми, которые связаны с его работой лишь опосредованно и диктуются нашими целями и опытом. Тогда мы можем, например, в качестве p_i взять стратегии вычисления, рассматривая переход от одной стратегии к другой в терминах логической системы, т.е. истолковывая этот переход как связь между формулами-стратегиями в рамках логической системы, обладающей определенной спецификой отношения следования, которая, как мы уже видели, сказывается на всей конструкции интенциональности. Похоже, логическая система играет в подобных конструкциях роль контекста, а, согласно Деннету, значение всегда зависит от всего контекста, следовательно, притворность нашего полагания как раз и есть выбор логической системы.

Вернемся теперь к пропозициональному содержанию p_i , приписываемому паре $\langle N_i, S_i \rangle$, которое осталось вне сферы нашего

¹ Блинов А. Общение. Звуки. Смысл. М., 1996.

внимания. Если пропозициональное содержание должно каким-то образом отображать реальность, то в качестве логической системы для описания подобных аспектов можно было бы выбрать систему формальной онтологии или онтологии, например, систему Онтологии Лесьневского. Подобные системы характеризуются тем, что их теоремы являются, с одной стороны, логическими утверждениями, а с другой стороны – онтологическими положениями. Если пропозициональное содержание будет представлено формулами системы этого типа, то мы в какой-то степени выполним требования, предъявляемые к общей структуре интенциональности.

Ограничивается ли применение интенциональной установки именно таким уровнем работы компьютера? Иначе говоря, должны ли «физические» состояния быть действительно физическими, обязательно соответствующими аппаратному уровню работы компьютера? Что будет, если мы сдвинем нашу установку на один уровень вверх, отказавшись от требования каузальности связи физических состояний? В данном случае стоит напомнить, что каузальную логику, в которой непосредственно бы описывалась каузальная связь (например, как некая каузальная импликация), до сих пор построить не удалось. В этом смысле отказ от описания каузальности связи физических состояний вполне мотивирован.

Сдвигая установку, в качестве «физических» состояний можно рассматривать вычислительные состояния, т.е. считать вычисления, производимые компьютером, аналогом физических состояний нейрона. Под этим углом зрения можно было бы рассматривать работу прувера, доказывающего теоремы. Логикой этих состояний тогда, возможно, например, будет модальная логика доказуемости. Возникает вопрос: необходимо ли вообще принимать в качестве логики «ментальных» состояний классическую двужначную логику? Ведь понятие «ментальность» теперь тоже можно сделать относительным, отказываясь от жесткой привязки к классической логике, как обязательной логике сознания искусственного интеллекта. И «физичность» и «ментальность» были связаны с решением задачи подхода к сознанию в рамках логического функционализма, однако теперь перед нами стоит совершенно иная задача.

Заметим, что ничто не препятствует нам, в частности, рассматривать Онтологию Лесьневского в качестве «ментальной» логики, поскольку она позволяет получать онтологические утверждения логическим путем. Проблемы на этом пути возникают (как и в первом

случае) из-за ряда особенностей этой системы, которые выводят ее за рамки стандартного логического формализма, приводя к ненужным осложнениям. Но, во-первых, действуя в этом направлении, можно воспользоваться и иными системами онтологии, более тесно связанными с такого рода формализмом. Во-вторых, если буквально истолковывать интенциональность установки, то можно, например, рассмотреть в качестве логики подобной интенциональности модальные логики знания и веры. В этом случае состояния компьютера описываются выражениями типа «известно, что в мире имеет место то-то и то-то положение дел, основанное на доказательстве такого-то логического утверждения».

Но и это не конец: можно использовать в качестве логики интенциональности не-фрегевскую логику, когда речь идет о корелативности формул, т.е. о совпадении ситуаций, описываемых этими формулами. Здесь возникают описания типа «такие-то и такие-то ситуации совпадают при указанном положении дел, связанном с доказательством данного логического утверждения». Таким образом, используя различные неклассические логики, мы получаем соответствующие различные описания.

Для реализации интенциональной установки этого и других уровней нам фактически не требуется никаких иерархических, реляционных конструкций или нейронных цепей. Всегда можно сконструировать логику, которая будет эквивалентна произведению трех логик, т.е. логике троек $\langle N_i, S_i, p_i \rangle$, которую и следует применить для описания системы искусственного интеллекта. Конкретный вид интенциональности дается тогда обратным разложением формул на тройки формул и тройки соответствующих выводов.

В качестве последнего контраргумента против витгенштейнианского ИИИ можно рассматривать проблему природы неклассических логических систем, используемых в его конструкции. Дело в том, что при рассмотрении вторичной интенциональности с помощью троек $\langle N_i, S_i, p_i \rangle$ неясным моментом остается логический статус третьего элемента N_i , представляющего собой физические состояния нейрона. Что это может быть за логика, описывающая функционирование нейронов? Точнее, можем ли мы вообще использовать логику с этой целью?

Здесь на помощь приходит понятие постовских систем, иначе, индуктивных определений, или дедуктивных систем (в терминологии С. Ю. Маслова), которое было введено в 1943 г. американским

логиком польского происхождения Э. Постом. Пост заметил, что в процессе наших рассуждений мы можем сосредоточиться на самих правилах вывода одних высказываний из других, не обращая никакого внимания на вид этих высказываний. Тогда правила вывода рассматриваются просто как некие правила получения одних абстрактных объектов из других в рамках абстрактной «квазилогической» системы.

Такие системы не обязательно должны быть тенью систем логических рассуждений, и могут жить собственной жизнью, движимые своими внутренними потребностями. Сам общий класс таких определений можно изучать, абстрагируясь от конкретного типа формальных систем. Например, можно рассматривать множества всех возможных эволюционных потомков фиксированного начального индивида J (считая, что индивиды полностью определены своим генотипом, а генотипы – это слова в подходящем алфавите), где шагом эволюции будет рождение всякого нового плодovitого мутанта. Дедуктивная система, описывающая различные (ведущие к плодovitому мутанту) мутации, будет системой с аксиомой J и с некоторыми правилами вывода. Для случая эволюции слова ab при трех допустимых типах «мутаций» эти правила выглядят следующим образом: $p \rightarrow pp$ (полиплодия), $pbabQ \rightarrow pbbQ$ (делеция-1), $pabbQ \rightarrow rabQ$ (делеция-2).

Современные системы неклассической логики во многих случаях представляют собой именно подобные системы. Во всяком случае, ничто не мешает нам использовать дедуктивные системы для описания функционирования нейронов, да, собственно говоря, и каузального уровня витгенштейнианского ИИ.

ЛИТЕРАТУРА

1. Брентано Ф. Избранные работы, М., 1996.
2. Блинов А. Общение. Звуки. Смысл. М., 1996.
3. Васюков В. Л. Формальная онтология. М., 2006.
4. Васюков В. Л. Проблема контекста интерпретации в универсальной логике // Логические исследования. Вып. 14. М., 2007. С. 105-130.
5. Витгенштейн Л. Логико-философский трактат. М.: Канон+, 2008.
6. Маслов С. Ю. Теория дедуктивных систем и ее применения. М.: Радио и связь, 1986.
7. Мерло-Понти М. Феноменология восприятия. СПб.: «Ювента» «Наука» 1999.
8. Сёрль Дж. Р. Природа интенциональных состояний // Философия. Логика. Язык. М., 1987. С. 96-126.

9. Сёрль Дж. Р. Рациональность в действии. М., 2004.
10. Смирнов В. А. Логический анализ научного знания М.: УРСС, 2001.
11. Dreyfus H. L. Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian // Philosophical Psychology 20 (2), 2007.
12. Freeman W. Societies of Brains: A study in the neuroscience of love and hate. The Spinoza Lectures. Amsterdam: Lawrence Erlbaum Associates, 1995.
13. Post E. Formal reductions of the general combinatorial decision problem // Amer. J. Math., vol. 65, 1943. P. 197-215.
14. Searle J. R. The Future of Philosophy (Article for the millenium proceedings of the Royal Society) // URL: http://socrates.berkeley.edu/~jsearle/future_of_philosophy.rtf
15. Searle J. R. Philosophy in a New Century. N. Y. : Cambridge University Center, 2008.

КОМПЬЮТЕРНЫЕ МОДЕЛИ ВРЕМЕНИ COMPUTER MODELS OF TIME

阿泥所扶 «时间的电脑模型»

**Анисов
Александр Михайлович**

*доктор философских наук, профессор, ведущий научный
сотрудник Института философии РАН, г. Москва*

Anisov Alexander

*Doctor of Philosophy, Professor, Leading Researcher,
Institute of Philosophy, RAS, Moscow*



Аннотация. Некоторые свойства времени позволяют представлять их геометрическими, т.е. пространственными средствами. Но было бы ошибочным сводить феномен времени к пространству, как это фактически делается в современной физике с её теорией единого пространства-времени. Если попытаться выразить суть времени в немногих словах, то можно сказать так: время – это вычислительный процесс. Переход от настоящего к прошлому и будущему – это результат вычислений, производимых самой природой. Поэтому в качестве базовых средств представления времени должны использоваться компьютерные модели. Имеющиеся в науке теории вычислимости здесь не очень-то пригодны и требуют обобщений, ориентированных именно на моделирование времени.

Ключевые слова: время, становление, компьютер, вычислимость, модель, вселенная, статика, динамика.

Annotation. Some properties of time may be conceived geometrically, i.e. envisioned with the help of spatial tools. But it would be incorrect to reduce the time phenomenon to space as it is actually assumed in modern physics and in the theory of unified time-space. Attempting to express the essence of time in a few words it is admissible to assert that time is a computation process. A transition from the present to the past and to the future is a result of calculations carried out by nature. Therefore computer models should be utilized as a basic means for presenting time. Unfortunately computation theories available in science today are mostly unsuitable for the purpose and demand far-reaching generalizations that concern modeling time in particular.

Keywords: Time, becoming, computer, computability, universe, statics, dynamics.

Процедуры Creation, Oblivion и Past

Острота проблемы становления связана, в первую очередь, с коллизией, возникшей между кажущимся несомненным свидетельством в пользу наличия феномена течения времени со стороны чувственного опыта и философской традиции, и столь же уверенным выводом об отсутствии течения времени, сделанным со ссылкой на авторитет физической науки. Мы присоединяемся к мнению К. Смита, следующим образом оценившего сложившуюся ситуацию: «Один из центральных спорных вопросов философии времени XX столетия – является ли временное становление независимой от мышления особенностью, присущей событиям, или кажимостью, присущей свойству человеческого понимания?» [14, p. 109].

Насколько нам известно, до сих пор в отечественной и мировой литературе не были рассмотрены точные конструкции, позволяющие реализовать хотя бы несложные модели течения времени или становления. В первых трёх параграфах данной статьи предлагается пример реализации простейшей модели становления. Этот пример ни в коей мере не претендует на учёт всех сторон темпоральности. Более того, сложность этого примера находится на таком предельно низком уровне, какой вообще можно было допустить, не боясь полностью утратить свойство адекватности модели. Следующая особенность предлагаемой модели связана с возможностью реализовать её на реально существующих компьютерах.

В компьютерной программе, выполнение которой моделирует течение времени, роль моментов времени фактически (в дальнейшем это будет уточнено) играют массивы, занимающие всю (или почти всю) доступную для размещения информации оперативную память. Точнее говоря, попытка объявить в программе ещё один массив должна приводить к ошибке, связанной с нехваткой памяти. Философский смысл данного требования в том, что течение времени изменяет всё, что может быть изменено. В ходе становления изменяются как сами моменты времени, так и составляющие их события. Размещение (или даже только возможность размещения) в оперативной памяти момента-массива, не вовлечённого в процесс становления, было бы подобно возникновению неподвижного слоя воды в быстро текущем потоке.

Если моменты времени представлены в модели при помощи массивов, то естественно считать элементы массивов событиями. Внутренняя структура событий сейчас неважна, поэтому будем в

качестве событий использовать целые положительные числа (число 0, по определению, событием не считается). Утверждение «В момент времени t произошло событие s » будет, таким образом, означать, что в памяти компьютера расположен массив t , содержащий число s .

Выполнение программы, написанной на BASIC, будет сводиться к вызову в определённом порядке трёх процедур в ходе бесконечного цикла. Рассмотрим эти процедуры.

Процедура Creation. Говоря неформально, эта процедура обеспечивает появление нового. Новые события входят в мир через настоящее. Процедура Creation и обеспечивает появление очередного нового настоящего. Её роль заключается в заполнении массива (всегда одного и того же) новыми событиями, для получения которых используется оператор RND, выдающий псевдослучайные целые числа. Предположим, оператор действует недетерминированным образом, так что предсказать появление конкретного числа невозможно в принципе. В реальных компьютерах это не так, однако технически возможно присоединить к компьютеру физически реализованный генератор случайных чисел.

Процедура Oblivion. Это процедура забвения. Прежде, чем при помощи процедуры Creation будет создано новое настоящее (при условии, что оно не первое), часть информации о событиях старого настоящего должна быть забыта. Почему часть, а не вся информация? По той причине, что настоящее в реальной действительности оставляет следы, не исчезает полностью в следующее мгновение. Предотвратить потерю части информации о настоящем невозможно, так как ресурсы пространства-памяти ограничены и нигде будет разместиться новое настоящее, идущее на смену прежнему.

Так же, как настоящее должно уступить место грядущему настоящему, «вчера» переходит в «позавчера», освобождая путь для нового «вчера», вновь теряя при этом информацию о событиях благодаря процедуре Oblivion. Вообще переход любого прошлого момента времени в более далёкое прошлое сопровождается потерями, определяемыми этой процедурой. Так что процедуре Oblivion подвержены в итоге все моменты – от настоящего до самого далёкого прошлого. При этом, поскольку ресурсы памяти ограничены (память реальных компьютеров конечна), самый ранний момент прошлого, находящийся в памяти, на очередном шаге становления не может быть сохранен даже частично. Данный момент должен исчезнуть

полностью. Память о событиях, случившихся в этот момент, оказывается утраченной навсегда.

С технической точки зрения действие процедуры Oblivion состоит в обнулении ненулевых элементов обрабатываемого массива, выбранных при помощи оператора RND. Процедура Oblivion сначала обрабатывает массивы с меньшим объёмом занимаемой памяти, а затем переходит на массивы большей величины. Как мы увидим, такой порядок соответствует темпоральному отношению «раньше, чем», т.е. потеря информации происходит вначале в более ранних моментах времени, а затем в более поздних. Информация из настоящего теряется последней.

Процедура Past. Назначение процедуры состоит в пересылке сохранившегося после работы процедуры Oblivion объёма информации в прошлое. Так, если массив-момент t является настоящим, а затем с помощью Oblivion теряет часть информации, то оставшаяся информация должна уйти в прошлое, освободив место в массиве t для новой информации, получаемой при использовании процедуры Creation. Если ближайшее прошлое это массив t^* , то процедура Past пересылает сохранившиеся события из массива-момента t в массив-момент t^* . В результате образуется момент прошлого t^* . В последующем сохранившаяся информация из t^* должна быть отправлена в массив-момент t^{**} – более далёкое прошлое для t , чем t^* , и т.д.

Мы видим, что незамысловатые рассуждения о том, что будущее становится настоящим, настоящее становится прошлым, а прошлое переходит в ещё более далёкое прошлое, явно недостаточны при описании течения времени. Становление – сложный процесс, включающий в себя ряд этапов, без которых не обойтись даже в наиболее простых моделях течения времени. Но было бы крайне непоследовательным считать, что эти этапы протекают во времени. Они и есть время. Укажем в этой связи на неправомерность присвоения звания моментов времени любым шагам процесса становления. Это не моменты времени, а этапы создания моментов времени в ходе течения времени. Последовательность этапов становления не есть последовательность моментов времени или последовательность мгновений.

Поэтому, например, нельзя ставить вопросы о том, как быстро течёт время, какова скорость становления и т.п. Такие вопросы обнаруживают полное непонимание сути дела. Никакого «внешнего» времени, по отношению к которому можно было бы оценивать темп становления, в модели нет, его там не существует вовсе. Конечно, мы

можем расширить темпоральный универсум при помощи введения внешнего времени. И тогда по этому внешнему времени оценивать скорость протекания вышеописанных процедур. Но это будет другой универсум, который для нас не важен по той причине, что не имеет никакого отношения к моделированию становления. Ведь в реальном изменяющемся во времени универсуме нет никаких признаков внешнего по отношению к нему метавремени. Нельзя же за такие признаки принимать фантастические сочинения о мультиверсе, параллельных вселенных и тому подобных продуктах не в меру разывгравшегося геометрического воображения.

Программа BECOMING

Рассмотрим написанную на языке BASIC программу BECOMING, моделирующую течение времени. Приведём вначале текст самой программы.

```
1 REM Программа BECOMING
5 X1%=38 : X2%=800 : X3%=2000 : X4%=4000
8 DIM V1%(X1%) : DIM V2%(X2%) :
  DIM V3%(X3%) : DIM V4%(X4%)
10 PROCCreation
20 PROCObivion(«V4%»)
30 PROCPast(«V3%»)
40 PROCCreation
50 PROCObivion(«V3%»)
60 PROCObivion(«V4%»)
70 PROCPast(«V2%»)
80 PROCPast(«V3%»)
90 PROCCreation
100 PROCObivion(«V2%»)
110 PROCObivion(«V3%»)
120 PROCObivion(«V4%»)
130 PROCPast(«V1%»)
140 PROCPast(«V2%»)
150 PROCPast(«V3%»)
160 PROCCreation
200 REPEAT
210 PROCObivion(«V1%»)
220 PROCObivion(«V2%»)
230 PROCObivion(«V3%»)
240 PROCObivion(«V4%»)
250 PROCPast(«V1%»)
260 PROCPast(«V2%»)
270 PROCPast(«V3%»)
280 PROCCreation
290 PRINT «METAMOMENT»
300 UNTIL FALSE
310 END
320 :
330 DEF PROCcreation
340 FOR I%=1 TO X4%
350 V4%(I%)=RND(1000000000)
360 NEXT I%
```

```
370 ENDPROC
380 :
390 DEF PROCObivion(X$)
400 C%=0 : E%=0
410 IF X$=»V1%»
    FOR I%=1 TO X1% : V1%(I%)=0 :
    NEXT I% : GOTO570
420 IF X$=»V2%» GOTO 450
430 IF X$=»V3%» GOTO 490
440 IF X$=»V4%» GOTO 530
450 REPEAT
460 E%=RND(X2%)
470 IF V2%(E%)>0 V2%(E%)=0 : C%=C%+1
480 UNTIL C%=X2%-X1%
485 GOTO 570
490 REPEAT
500 E%=RND(X3%)
510 IF V3%(E%)>0 V3%(E%)=0 : C%=C%+1
520 UNTIL C%=X3%-X2%
525 GOTO 570
530 REPEAT
540 E%=RND(X4%)
550 IF V4%(E%)>0 V4%(E%)=0 : C%=C%+1
560 UNTIL C%=X4%-X3%
570 ENDPROC
580 :
590 DEF PROCPast(X$)
600 C%=0
610 IF X$=»V1%» GOTO 640
620 IF X$=»V2%» GOTO 700
630 IF X$=»V3%» GOTO 760
640 FOR I%=1 TO X1%
650 REPEAT
660 C%=C%+1
670 UNTIL V2%(C%)>0
680 V1%(I%)=V2%(C%)
690 NEXT I%
695 GOTO 820
700 FOR I%=1 TO X2%
710 REPEAT
720 C%=C%+1
730 UNTIL V3%(C%)>0
740 V2%(I%)=V3%(C%)
750 NEXT I%
755 GOTO 820
760 FOR I%=1 TO X3%
770 REPEAT
780 C%=C%+1
790 UNTIL V4%(C%)>0
800 V3%(I%)=V4%(C%)
810 NEXT I%
820 ENDPROC
```

Реально программа BECOMING выполнялась на персональном компьютере BBC Acorn. Это был один из первых персональных компьютеров, не имевший жёсткого диска, с ничтожной по современным меркам оперативной памятью и единственным языком высокого

уровня BASIC. Для других компьютеров язык программирования, размеры и количество массивов могут быть другими. Важно только, чтобы два массива, способные представлять момент настоящего, нельзя было одновременно разместить в области доступной оперативной памяти. Это требование философского порядка, так как никаким разумным образом невозможно (в рамках динамической концепции времени) оправдать возможность сосуществования двух равноправных «настоящих».

Более того, вообще попытка добавления ещё одного массива к n уже имеющимся массивам-моментам должна приводить к ошибке, связанной с нехваткой памяти при выполнении программы типа BECOMING. В противном случае наличие лишь n массивов-моментов нельзя содержательно оправдать. Кроме того, массивы должны иметь попарно различные размеры, так как в соответствии с общим принципом динамической концепции времени переход в прошлое связан с потерей информации. Следовательно, массивы, хранящие информацию о более далёком прошлом, должны быть меньшего объёма, чем массивы, хранящие информацию о менее далёком прошлом.

Выполнение программы BECOMING порождает бесконечный цикл. Никаких «внутренних» способов его остановить нет. Хотя моментов времени в нашем случае всего четыре, возникают они неограниченное число раз. С философской точки зрения это означает, что становление может длиться гораздо дольше, чем это отражено в существующей в универсуме памяти о прошлом. Так, никаких следов возрастом более полутора десятков миллиардов лет в нашем реальном универсуме не находят. Но это, вообще говоря, не ведёт к выводу, что он существует именно указанное количество лет. Возможно, даже если никаких более древних следов в действительности нет, его существование в циклах становления было существенно более длительным.

Выполнение строк с 10 по 160 однозначно соответствует шагам этапа инициализации множества моментов времени. Так, шаг 1 соответствует строке 10 программы, шаг 2 – строке 20 и т. д., т.е. шаг i диаграммы соответствует строке $10 \times i$ программы BECOMING, вплоть до шага $i = 16$ включительно.

Этап инициализации, или, проще говоря, создания множества моментов модельного времени, необходим по следующей причине. После выполнения строки 5 память компьютера будет распределена

между массивами V1%, V2%, V3% и V4%, которые, однако, мы не можем считать моментами времени. Ведь, по определению, моментом времени называется множество событий, а ни один из перечисленных массивов ещё не содержит ни одного события, т.е. целого числа $s > 0$, размещённого в массиве в результате выполнения соответствующих команд программы BECOMING.

Ясно, что массивы подлежат заполнению событиями. Но как это сделать? Если сначала превратить массив V4% в момент времени, а затем наполнить событиями массив V3%, то получится, что момент, содержащийся в V4%, наступил раньше, чем момент, содержащийся в V3%. Однако предполагалось, что наибольший по размерам массив V4% хранит события настоящего. Но настоящее не может наступить раньше прошлого. Если, с другой стороны, сначала заполнить V3%, то (с содержательной точки зрения) именно V3% и является моментом настоящего времени, поскольку других моментов пока просто нет! Не можем же мы, не впадая в противоречие со здравым смыслом, с самого начала объявить события из V3% прошлыми. Ведь тогда получилось бы, что существует прошлое, никогда не бывшее настоящим.

К счастью, можно избежать создавшихся трудностей. Путь, ведущий к выходу из тупика, соответствует этапу инициализации множества моментов, содержащей 16 шагов. Проследим за этими шагами (попутно указывая, с помощью каких процедур программы BECOMING осуществляются эти шаги).

На первом шаге вызывается процедура Creation (строка 10 программы), определение которой занимает строки с 330 по 370. Работая в цикле, процедура полностью заполняет массив V4% событиями, т.е. ненулевыми числами, выбираемыми оператором RND. На этом шаге после его завершения существует один-единственный момент – момент настоящего, содержащийся в массиве V4%. Обозначим его как $v(1)$. Никаких других моментов ещё нет, прошлого нет, отношение «раньше, чем» не определено.

На втором шаге не создаётся ещё какой-либо момент времени, – как мы видели, это привело бы к парадоксам, – а уничтожается часть событий момента настоящего при помощи вызова процедуры Oblivion (см. строку 20) с параметром $X\$ = \text{«V4%»}$, т.е. в качестве параметра берётся имя массива V4%, подлежащего частичному (в данном случае) обнулению. Недетерминированный выбор уничтожаемых событий имитируется оператором RND. После стирания части

событий настоящего уже нет. Но и прошлого в собственном смысле ещё нет – здесь фактически промежуточный этап создания прошлого.

Шаг 3 завершает создание первого момента прошлого при помощи вызова процедуры *Past* с именем массива *V3%* в качестве параметра (строка 30). В ходе выполнения процедуры *Past* все спасшиеся от уничтожения события бывшего настоящего пересылаются в массив *V3%*, который заполняется этими событиями целиком. В результате шаг 3 завершается созданием момента прошлого, содержащегося в массиве *V3%*. Момент из *V3%* – действительно момент прошлого, так как все события, составляющие этот момент, когда-то были событиями настоящего. Вместе с тем никаких других моментов нет: настоящее перестало быть настоящим, «провалилось» в прошлое, а массивы *V1%* и *V2%* вообще не содержат ни одного события, поэтому не содержат и моментов времени.

Но массив *V4%* содержит события! Должен ли он считаться моментом времени? Вроде бы раз массив содержит события, его следует отнести к моментам. Но, строго говоря, моментами времени мы называем множества событий. В конце концов не столь существенно, что эти множества в программе *BECOMING* представлены массивами. Важнее то, что с теоретико-множественной точки зрения множество событий в массиве *V4%* и множество событий в массиве *V3%* равны. Тем самым как множества событий массивы *V4%* и *V3%* совпадают. Они как бы склеиваются в одно множество. Если бы существовало хоть одно событие, принадлежащее *V4%*, но не принадлежащее *V3%* или наоборот, то мы имели бы дело с разными множествами событий и тем самым – с разными моментами времени. Но чего нет, того нет: на шаге 3 существует лишь одно множество событий, являющееся единственным моментом прошлого при отсутствии настоящего.

На шаге 4 впервые возникают два момента времени (обозначим их через $v(1)$ и $v(2)$). Появляется возможность определить отношение «раньше, чем», ответив на вопрос о последовательности возникновений $v(1)$ и $v(2)$. На этом шаге вновь вызывается процедура *Creation*, заново создающая настоящее. Поскольку момент *V3%* уже был в памяти перед тем, как появился новый момент *V4%*, естественно считать, что *V3%* раньше, чем *V4%*. Соответственно $v(1)$ раньше, чем $v(2)$. Действительно, $v(1)$ – это то, что осталось от предыдущего настоящего, тогда как $v(2)$ – вновь созданное настоящее. Ясно, что $v(1)$ принадлежит прошлому и предшествует во времени моменту $v(2)$.

На шаге 5 более ранний момент времени подвергается частичному уничтожению. Второй момент остаётся неизменным. С интуитивной точки зрения частично стёртый момент по-прежнему предшествует во времени моменту, сохранившемуся неизменным, так как события «нестабильного» момента появились в памяти компьютера раньше, чем события «стабильного» момента. Но нестабильность момента означает, что идёт процесс его трансформации в некий другой момент. Момент как бы перестал быть самим собой, но другим моментом не стал. В некотором смысле подвергшийся преобразованиям момент и момент, сохраняющий свою стабильность на очередном шаге становления, относятся к категориям разного рода. Так или иначе, мы будем определять модельное временное отношение «раньше, чем» только на «стабильных» моментах. Ситуации стабильности возникают на этапе инициализации или роста множества моментов на шагах 1, 4, 9 и 16, а на циклическом этапе, как мы увидим дальше, на шаге 0. Лишь на этих шагах (за исключением шага 1, когда имеется единственный момент времени) будет определено отношение R , являющееся модельным аналогом темпорального отношения «раньше, чем».

На шаге 6 частичному уничтожению подвергается момент настоящего (аналогично тому, как это происходило на шаге 2).

Шаг 7 состоит в переносе событий из второго по величине массива в третий по величине массив, который заполняется целиком и представляет впервые образованный ещё один момент прошлого.

Шаг 8 аналогичен шагу 7, за исключением того, что перенос событий осуществляется из наибольшего массива во второй по величине массив.

Шаг 9. На этом шаге впервые завершается создание трёх «полноценных» моментов времени, и, как уже говорилось, между ними устанавливается модельное отношение «раньше, чем». Ясно, что должно быть $v(1) R v(2) R v(3)$, так как события из $v(1)$ появились в памяти раньше всех, затем появились события из $v(2)$ и лишь впоследствии события из $v(3)$.

Шаги с 10 по 16 полностью аналогичны предыдущим. Отличие заключается только в том, что эти шаги приводят к образованию четырёх моментов времени, упорядоченных в соответствии со следующим порядком появления в памяти: $v(1) R v(2) R v(3) R v(4)$.

На циклическом этапе программа BECOMING входит в бесконечное повторение восьми шагов преобразований временного ряда.

Количество моментов более не возрастает из-за нехватки памяти (моделирующей ограниченность ресурсов в реальном универсуме).

Шаг 0. После выполнения шага 16 управление передаётся на шаг 1 цикла. Это означает, что перед выполнением первого шага цикла в памяти расположен «полноценный» ряд из 4 моментов, т.е. ряд, аналогичный тому, который должен появиться на шаге 0. Можно считать поэтому, что выполнение цикла начинается с шага 0, роль которого первоначально играет шаг 16, завершающий этап инициализации.

На шаге 1 самый ранний момент времени с помощью процедуры Oblivion подвергается полному уничтожению (массив V1% обнуляется). Полному – потому что сохранившуюся в первом моменте информацию невозможно (даже частично!) переслать в более далёкое прошлое из-за нехватки памяти. Между тем требуется место для новой информации.

Шаги 2 – 7 ничем не отличаются от шагов 10 – 15 диаграммы инициализации. В программе BECOMING этим шагам соответствует одна и та же последовательность процедур, вызываемых с теми же самыми параметрами.

Временные понятия в компьютерной среде

В чем смысл программы BECOMING? Она моделирует течение времени или становление. Введём ряд новых понятий, относящихся к программе BECOMING. Предположим, выполнение BECOMING находится на таком этапе, когда осмысленным становится вопрос о значениях, присвоенных элементам объявленных в BECOMING четырёх массивов. Будем говорить в этом случае, что массивы заполнены.

Каждый заполненный массив из множества $ARRAY =_{\text{Df}} \{V1\%, V2\%, V3\%, V4\%\}$ будем рассматривать как функцию из множества $\{1, 2, \dots, n\}$ (где n уменьшенная на 1 размерность массива) в некоторое подмножество множества $V =_{\text{Df}} \{0, 1, \dots, 10^9\}$. Обозначая через $\text{dom}(f)$ и $\text{rng}(f)$ области определения и значения функции f соответственно, положим

$$RNG =_{\text{Df}} \{m \mid \exists v (v \in ARRAY \ \& \ \text{rng}(v) = m)\}.$$

Назовём множество m из RNG квазимоментом (времени), если 0 принадлежит m и найдётся число, отличное от 0, также принадлежащее m :

$$\text{Квазимомент}(m) \leftrightarrow_{\text{Df}} (m \in RNG \ \& \ 0 \in m \ \& \ \exists x (x > 0 \ \& \ x \in m)).$$

Содержательно, множество m является квазимоментом, если оно содержит какие-либо события (роль которых играют отличные от нуля целые положительные числа), но может быть расширено при

помощи добавления других событий путём замены нулевых значений соответствующего массива на ненулевые.

Множество m из RNG является моментом (времени), если $0 \notin m$. Это эквивалентно утверждению $\forall x(x \in m \rightarrow x > 0)$. Итак,

Момент(m) $\leftrightarrow_{\text{Df}} (m \in \text{RNG} \ \& \ \forall x(x \in m \rightarrow x > 0))$.

Содержательно, множество m является моментом, если оно не может быть расширено при помощи добавления новых событий посредством замены нулевых значений соответствующего массива ненулевыми из-за того, что все элементы этого массива имеют значения, отличные от 0.

Множество всех квазимоментов и всех моментов обозначим через K и через M соответственно. Из определений немедленно следует, что $K \cap M = \emptyset$. На множестве $K \cup M$ определим отношение «квазираньше, чем», обозначив его через KR : $m KR m^*$, если и только если вызов процедуры, сформировавшей m , предшествовал вызову процедуры, сформировавшей m^* . «Нормальное» отношение R , соответствующее привычному «раньше, чем», является сужением отношения KR на множество моментов M : $m R m^* \leftrightarrow_{\text{Df}} (m \in M \ \& \ m^* \in M \ \& \ m KR m^*)$.

Введём ещё два важных определения. Назовём упорядоченную пару $KW =_{\text{Df}} \langle K \cup M, KR \rangle$ **квазиметамоментом**, если $K \neq \emptyset$. Упорядоченная пара $W =_{\text{Df}} \langle K \cup M, R \rangle$ является **метамоментом**, если $K = \emptyset$ и $M \neq \emptyset$.

В соответствии с этими определениями, метамоментам соответствуют на этапе инициализации шаги 1, 4, 9 и 16, а на циклическом этапе – шаги 1 и 0. Все остальные шаги обоих этапов порождают квазиметамоменты.

Существенно, что на каждом шаге становления возникает квазиметамомент или метамомент и притом только один. Два различных метамомента или два различных квазиметамомента сосуществовать вместе не могут – для их размещения не хватит памяти компьютера. Памяти хватает для размещения одного и только одного метамомента или квазиметамомента.

Предположим, что процесс выполнения программы BECOMING находится на таком этапе, когда в памяти компьютера размещён метамомент M . **Моментом настоящего** назовём момент $h \in M$, удовлетворяющий условию $(\forall m \in M)(m R h \vee m = h)$.

Множество моментов прошлого P образуем при помощи определения $P =_{\text{Df}} \{m \mid m \in M \ \& \ m R h\}$.

Казалось бы, настала очередь моментов будущего. Однако в рассматриваемой модели моменты будущего отсутствуют. Ведь момент – это множество событий (считающихся одновременными в рамках момента). Появление в памяти событий «будущего» означает, что эти события уже произошли. Тем самым к таким событиям применима оценка «уже», которая характерна как раз для событий, не являющихся будущими, – для событий прошлого и настоящего. Это прошлое уже произошло, и это настоящее уже есть. Не скажешь ведь, что будущее «уже будет». Как тут не вспомнить А. Бергсона, предложившего концепцию времени без будущего [4]. Однако просто сказать, что будущего нет и этим ограничиться – значит дать ввести себя в заблуждение. В определённом смысле будущее существует, но это существование нельзя трактовать по типу прошлого и настоящего.

Теперь мы можем приступить к опровержению доводов Мак-Таггарта об иллюзорности времени [12], используя для их анализа только что созданный новый понятийный аппарат. Допустим, событие s есть настоящее, т.е. $s \in h$. Означает ли это, что s было будущим и будет прошлым? «Было будущим», согласно Мак-Таггарту, равнозначно «является будущим в некоторый момент прошлого». Пусть этим моментом прошлого является $m \in P$. Можно ли утверждать, что событие s является будущим в момент m ? Нет, поскольку будущее – это то, что наступает вслед за настоящим, а m – момент прошлого. Но ведь m когда-то был моментом настоящего? Опять-таки нет, потому что m это осколок, след исчезнувшего настоящего. И когда m входил составной частью в ушедшее настоящее, появление события s отнюдь не было предопределено: в будущем могло случиться так, что $m \in P$ и $s \notin h$. Равно как могло случиться $m \in P$ и $s \in h$, $m \notin P$ и $s \notin h$ или $m \notin P$ и $s \in h$. Произошло так, что $m \in P$ и $s \in h$, но это дело случая, реализовавшегося в момент «теперь», а не предуготовленного будущим. Во времена диктатора Гая Юлия Цезаря просто не существовало времени, в котором пребывала королева Анна Стюарт, и, значит, событие её смерти не было будущим. В те времена было неизбежным событие смерти Цезаря, но и тут будущее не было полностью фиксированным: заговор мог быть раскрыт, Цезарь мог умереть от внезапной болезни и т.д.

Далее, «будет прошлым», по Мак-Таггарту, эквивалентно «является прошлым в некоторый момент будущего». Применительно к рассматриваемому примеру получается, что в будущем, когда момент настоящего h перейдёт в прошлое, событие s станет событием

прошлого. Однако моментом прошлого станет не момент h , а какая-то его часть. И вновь ничем не предопределено, что событие s останется в этой части. Как знать, быть может, когда-нибудь в веках на чисто потеряются следы существования диктатора Юлия Цезаря и королевы Анны Стюарт.

Стало быть, как то, что событие настоящего s было будущим, так и то, что s будет прошлым, весьма проблематично. Так что же, доводы Мак-Таггарта об иллюзорности времени следует признать ошибочными? В целом, да. Но Мак-Таггарт прав частично: его рассуждения показывают, что такие темпоральные характеристики, как прошлое, настоящее и будущее, не могут быть заданы при помощи рядов, являющихся, по сути, геометрическими образованиями. Когда только что шла речь о проблематичности судьбы события s или события смерти королевы Анны, нами имелся в виду негеометрический, программный аспект реальности, который невозможно представить комбинацией геометрических рядов.

Существует совершенно не поддающаяся учёту в геометрических построениях асимметрия между осуществившимся прошлым и настоящим и неосуществившимся будущим: тот факт, что настоящее и прошлое обрело бытие, не означает, что тем самым гарантирован и переход в бытие будущего. Ведь *в компьютерных моделях будущее может наступить или не наступить*. В случае останова программы вычислений течение модельного времени прекратится и будущее не наступит.

Можно было бы попытаться найти оправдание программным образом реализованному завершению модельного становления в том, что реальные компьютеры являются автоматами с конечным числом возможных состояний, в силу чего рано или поздно процесс течения модельного времени начнёт повторять сам себя. Иначе говоря, программы, моделирующие становление на реальных ЭВМ посредством бесконечных циклов, неизбежно осуществляют в той или иной форме реализацию идеи *вечного возвращения*.

Однако остаётся непонятным, почему требуется заботиться о предотвращении вечного возвращения. В конце концов, вопрос о допустимости вечного возвращения относится к числу концептуальных и предрешать его решение по формально-техническим соображениям нет оснований. Напротив, с технической точки зрения моделировать переход от прошлого к будущему удобнее при помощи бесконечных циклов, что влечёт осуществление вечного возвращения. Так

что вопрос скорее в том, насколько важны философские аргументы против вечного возвращения и достаточен ли их вес для того, чтобы изыскивать средства борьбы с технически совершенными бесконечными циклами в программах, моделирующих становление.

Небезынтересно отметить следующее: на реальных ЭВМ можно организовать потенциально бесконечный цикл, направленный в будущее, но нельзя построить потенциально бесконечный цикл, направленный в прошлое. С одной стороны, сколько бы шагов ни сделал компьютер, выполняя произвольную программу α число этих шагов остаётся конечным. С другой стороны, может случиться так, что число шагов по выполнению программы α в будущем неограничено. Первая особенность реальных компьютерных моделей времени делает неизбежным существование начала времени, вторая обуславливает вечное возвращение при неограниченном определённым числом повторении ответственной за возникновение циклов части команд программы α .

Ограничения, присущие реализованным на действительно существующих ЭВМ моделям становления, могут быть преодолены только в абстракции. Перейдём к обсуждению абстрактных программ, моделирующих течение времени и выполняющихся на идеальных вычислительных устройствах.

Абстрактная вычислимость и язык программирования АВТ

Статическая модель мира хорошо «схватывается» метафорой Книга Природы (Г.Галилей и др.). Эта модель не оставалась неизменной и эволюционировала от первоначальных попыток вместить реальность в механическую картину до вполне современных и принятых в науке геометрических моделей мира (например, неевклидово пространство-время теории относительности). В последнее время, однако, намечается тенденция, которую можно выразить метафорой *Компьютер Природы*. В связи с осмыслением тех интеллектуальных изменений, которые принесла с собой компьютерная революция, начинают появляться ростки негеометрических идей в описании физической реальности ([5], [8], [9], [13], [15]). Однако не стоит преувеличивать их значение: реальных угроз господствующей геометрической парадигме в точном естествознании они пока не представляют.

Предпринимаются усилия по созданию обобщённых теорий вычислимости. К их числу относится, например, рекурсия в высших типах [6], теория α -рекурсии, позволяющая выходить за границы натурального ряда ω и вычислять до допустимого ординала $\alpha > \omega$ [10],

или теория машин Тьюринга, способных вычислять за трансфинитное время [11]. Пока эти и им подобные обобщения находятся вне рамок задачи построения компьютерных моделей универсума, но, возможно, это временное положение дел и будущие усовершенствованные теории обобщённой вычислимости окажутся более адекватными для этой цели. Кроме того (и это главное), эти обобщения исходят из стремления получить аналог обычной теории вычислимости, и в этой связи упор делается на обобщение идеи *эффективности*. Но суть дела состоит в том, что не всякое обобщение идеи вычислимости удовлетворительно с концептуально-философской точки зрения. Мир, в котором мы существуем, является совокупностью разного рода процессов, большинство из которых трудно отнести к эффективно организованным. Значит, необходима *теория неэффективных процессов*, которые выходят за границы детерминизма.

Хотелось бы, кроме того, иметь такую теорию неэффективной вычислимости, в которой любой процесс *локально* вёл бы себя как обычный вычислительный процесс: процессы должны состоять из шагов, каждый из которых (если он не первый и не последний) выполняется при условии, что выполнен непосредственно предшествующий шаг, и что выполнение очередного шага вызывает осуществление непосредственно следующего шага. Между тем, в рамках имеющих обобщений понятия вычислимости допустимы, например, процессы, содержащие $\omega+1$ число шагов (ω – это натуральный ряд $0, 1, 2, \dots, n, \dots$, за которым следует трансфинитный шаг $\omega+1$). В качестве иллюстрации можно привести решение проблемы останова обычной вычислительной машины на обобщённой машине, которое потребует как раз $\omega+1$ шагов [7, с. 520]. Последний шаг при таком понимании налицо, однако нельзя указать тот конкретный шаг, осуществление которого непосредственно предшествовало выполнению последнего шага.

Автор предлагает в качестве альтернативы построенную им нестандартную теорию индетерминированной вычислимости – АВТ-вычислимости ([1], [2], [3]). В нынешнем виде теория АВТ-вычислимости, по-видимому, далека от совершенства. И всё-таки это пусть малый, но шаг вперёд в построении возможной будущей онтологии, реализующей метафору *Компьютера Природы*.

Рассмотрим множество, упорядоченное по типу $\omega+\omega^*$. Эту структуру можно представлять как натуральный ряд, к «концу» которого добавлен ряд отрицательных целых чисел. В отличие от $\omega+1$, в $\omega+\omega^*$

нет выделенных точек в том смысле, что любой элемент α из $\omega + \omega^*$, имеющий предшествующий элемент, имеет и непосредственного предшественника, а имеющий последующий элемент, имеет и непосредственного последователя.

Что произойдёт, если мы заменим все элементы из $\omega + \omega^*$ электрическими лампочками и зажжём первую лампочку, потом вторую, потом третью? Что будет? Интуитивно кажется, что мы будем всё время находиться где-то на каком-то n «внутри» ω . Это так, если эти лампочки зажигаются во времени. А если они сами представляют основания времени? Если это не процесс, протекающий во времени, а если это *само время*, которое так устроено, причём устроено на уровне, гораздо более глубоком, чем нам это дано эмпирически? В таком случае нет никакого противоречия в постулате, что этот процесс последовательного включения и выключения лампочек рано или поздно придёт к концу, и у нас будет гореть в итоге последняя лампочка. И вновь здесь не будет никаких трансфинитных скачков (подобных, например, переходу от ω к $\omega + 1$), то есть локально всё происходит совершенно обычным образом, как если бы лампочек было конечное число.

Теперь сделаем следующий шаг. Как известно, любая первопорядковая непротиворечивая формальная аксиоматическая теория, имеющая бесконечную модель, по теореме Лёвенгейма-Сколема-Тарского из математической логики, имеет модель *произвольной мощности*. Арифметика – как раз пример такой теории. Ведь натуральных чисел бесконечно много, а разумные люди не сомневаются в непротиворечивости арифметики. Тогда по теореме мы можем взять отрезок натурального ряда $0, 1, 2, \dots, n, \dots, N, \dots$, где N будет иметь мощность континуума. Заметьте, этот ряд подчиняется аксиомам арифметики, ничего тут больше нет, никаких других аксиом нет. Значит, ничто не мешает представлять этот ряд полученным *за несчётное число дискретных шагов* при помощи подходящего АВТ-компьютера. А раз числу N предшествует несчётное множество гипернатуральных чисел, мы можем организовать взаимно однозначное отображение ряда $0, \dots, N$ на отрезок обычного линейного континуума.

Будущее АВТ-вычислений способно индетерминированно ветвиться. В отличие от линейного прошлого (Р) и настоящего (Н), будущее (F) представлено множеством альтернатив. В итоге получается структура времени, совершенно непохожая на стрелу. Скорее, в графическом изображении она напоминает метлу. Царит сметающая

людей, цивилизации и звёздные миры *метла времени*! В этой связи предлагается устаревшую метафору Эддингтона о стреле времени заменить метафорой *метлы времени*. Стрела времени представляет статический геометрический универсум, в котором всё прописано и всё уже существует. Метла времени – образ динамического, компьютерного универсума, оставляющего место для индетерминированного выбора будущего.

В этом компьютерном универсуме возникают совершенно специфические логические вопросы. А не застрянет ли компьютерное течение времени? Ведь исполняемые компьютером программы обладают свойством зависеть не только из-за поломки компьютера, но также из-за логических несоответствий, которые могут быть в этих программах. Поэтому необходимо доказывать формальную корректность программ, моделирующих течение времени, на предмет исключения возможности логического авоста (аварийного останова).

На языке программирования АВТ можно формально точно выразить идеи дискретного пересчёта континуума и течения времени. Абстрактные АВТ-компьютеры в ходе выполнения АВТ-программ способны реализовывать дискретные трансфинитные процессы и размещать в своей памяти (если она бесконечна) некоторые бесконечные множества (если их мощность не превышает мощности свободной памяти). К сожалению, рамки статьи исключают возможность сформулировать здесь теорию АВТ-вычислимости, но её подробное описание можно найти в [1], [2] и [3].

Приведём примеры простейших АВТ-программ, иллюстрирующие особенности языка АВТ.

(π1)	(π2)
I_1 CHOOSE X X = T	I_1 CHOOSE X X = T
I_2 GOTO I_1	I_2 DELETE X
I_3 DELETE X	I_3 GOTO I_1
(π3)	(π4)
I_1 DELETE X	I_0 GOTO I_1
I_2 CHOOSE X X = T	I_1 CHOOSE X X = T
I_3 GOTO I_1	I_2 DELETE X
	I_3 GOTO I_1

Не вникая в технические детали, кратко охарактеризуем результаты попыток выполнения приведённых программ. Программа π1 успешно работать не будет в любом случае. Программа π2 может породить как имеющий начало бесконечный цикл вида $I_1, I_2, I_3, I_1, I_2, I_3, I_1, \dots$

, так и не имеющий ни начала, ни конца бесконечный цикл вида ..., $l_1, l_2, l_3, l_1, l_2, l_3, l_1, \dots$ (возможно также, что на каком-то шаге $\pi 2$ завершится аварийным остановом или авостом при попытке выполнить команду l_1). Программа $\pi 3$ может выполнять только не имеющие начала бесконечные циклы вида ..., $l_1, l_2, l_3, l_1, l_2, l_3, l_1, \dots$ или вида ..., $l_1, l_2, l_3, l_1, l_2, l_3, l_1, l_2$ (в последнем случае не имеющий начала цикл завершится авостом именно на команде l_2 ; возможно также, что $\pi 3$ не будет выполняться вообще). Наконец, программа $\pi 4$ может реализовывать только имеющие начало процессы. При этом они могут завершаться авостом, но могут не иметь конца: $l_0, l_1, l_2, l_3, l_1, l_2, l_3, l_1, \dots$.

Обсуждаемые программы важны для приложений в философии. Так, проблема начала времени не имеет устраивающего всех исследователей единственного решения. Если принимается тезис о том, что эта проблема неразрешима, то для моделирования течения времени больше подходит конструкция, аналогичная программе $\pi 2$; принятие тезиса об отсутствии начала течения времени заставит прибегнуть к программам типа $\pi 3$. Наконец, программы типа $\pi 4$ выражают идею начала времени. При этом компьютерное моделирование течения времени порождает специфическую проблему: попытка выполнения очередной команды может завершиться авостом. В этом случае модельное становление обрывается, АВТ-компьютер зависает, наступает «конец света».

АВТ-вычислимость позволяет моделировать течение времени. Предложенная нами теория **TS** есть не что иное, как формальное описание *метлы времени*. Аксиомами теории **TS** являются следующие формулы, образующие из-за пункта 1 бесконечный список.

1. $\forall X(m(X) \rightarrow \exists x_1 \exists x_2 \dots \exists x_n (X(x_1) \& X(x_2) \& \dots \& X(x_n) \& \neg(x_1 \approx x_2) \& \neg(x_1 \approx x_3) \& \dots \& \neg(x_1 \approx x_n) \& \neg(x_2 \approx x_3) \& \neg(x_2 \approx x_4) \& \dots \& \neg(x_2 \approx x_n) \& \dots \& \neg(x_j \approx x_{j+1}) \& \neg(x_j \approx x_{j+2}) \& \dots \& \neg(x_j \approx x_n) \& \dots \& \neg(x_{n-1} \approx x_n)))$
(для каждого $n > 1$ имеем отдельную аксиому 1_n)

2. $\forall x \exists X(m(X) \& X(x))$

3. $\forall X \neg(X \mathbf{E} X)$

4. $\forall X \forall Y \forall Z (X \mathbf{E} Y \& Y \mathbf{E} Z \rightarrow X \mathbf{E} Z)$

5. $\forall X(m(X) \rightarrow X || H)$

6. $\forall X \forall Y \forall Z (Y \mathbf{E} X \& Z \mathbf{E} X \rightarrow Y || Z)$

7. $\forall X(m(X) \rightarrow \exists Y(Y \mathbf{E} X \& Y || X))$

8. $\forall X(\exists Y(X \mathbf{E} Y) \rightarrow \exists Z(X \mathbf{E} Z \& X || Z))$

9. $\exists X \exists Y (H \mathbf{E} X \& H \mathbf{E} Y \& H || X \& H || Y \& X \neq Y)$

Метла времени не находится в фиксированном состоянии, а подвергается надлежащим преобразованиям посредством программы BECOMING.ABT, моделирующей течение времени на основе теории **TS**.

Программа BECOMING.ABT

```

1 DELETE X
2 CHOOSE X | (X отрезок LD) & X = Y*
3 IF X = LD THEN END
4 DELETE Y
5 CHOOSE Y | (Y отрезок LD) & Y = X*
6 IF Y = LD THEN END
7 DELETE X1
8 CHOOSE X1 | X1 |= TS & X1 – сужение в прошлое для Y1
9 DELETE Y1
10 CHOOSE Y1 | Y1 |= TS' & Y1 – переход в будущее относительно X1
11 DELETE X1
12 CHOOSE X1 | X1 |= TS & X1 – первое расширение в будущее для Y1
    & X1 корректен & |Mm(X)| ≤ |Mm\ (Mm(X) ∪ Mm(Y) ∪ Mm(X1) ∪ Mm(Y1))|
13 DELETE Y1
14 CHOOSE Y1 | Y1 |= TS & Y1 – второе расширение в будущее для X1
    & Y1 корректен & |Mm(X1)| ≤ |Mm(Y1)|
15 GOTO I1

```

Вопрос о корректности программы BECOMING.ABT весьма не-тривиален, но здесь он обсуждаться не будет.

Данная модель становления является универсумом с ограниченными ресурсами (*уором*). После выполнения инструкции I14 возможно, в частности, использование всей памяти без остатка для размещения сменяющих друг друга структур. Разумеется, предыдущая модель становления также *уор*. На примере построенных *уоров* хорошо видны, как нам представляется, специфические проблемы, возникающие в динамических универсумах, моделирующих течение времени.

ЛИТЕРАТУРА

1. Анисов А. М. Время и компьютер. Негеометрический образ времени. М., 1991.
2. Анисов А. М. Темпоральный универсум и его познание. М., 2000.
3. Анисов А. М. Феномен течения времени. Логико-философский анализ. – LAP LAMBERT Academic Publishing, 2012.
4. Бергсон А. Время и свобода воли. М., 1910.
5. Беркович С. Я. Клеточные автоматы как модель реальности: поиски новых представлений физических и информационных процессов. М., 1993.
6. Кекрис А., Московакис Я. Рекурсия в высших типах // Справочная книга по математической логике. Ч. III. Теория рекурсии. М., 1982.

7. Роджерс Х. Теория рекурсивных функций и эффективная вычислимость. М., 1972.
8. Тоффоли Т., Марголюс Н. Машины клеточных автоматов. М., 1991.
9. Хармут Х. Применение методов теории информации в физике. М., 1989.
10. Шор Р. Теория α -рекурсии // Справочная книга по математической логике. Ч. III. Теория рекурсии. М., 1982.
11. Hamkins J. D., Lewis A. Infinite Time Turing Machines // The Journal of Symbolic Logic, V. 65, N. 2, June 2000, P. 567 – 604.
12. McTaggart J. E. The Nature of Existence. Cambridge, 1927.
13. Poundstone W. The Recursive Universe. Cosmic Complexity and the Limits of Scientific Knowledge. N. Y., 1985.
14. Smith Q. The Mind-Independent of Temporal Becoming. //Philosophical Studies, Dordrecht, 1985, vol. 47, N 1.
15. Wolfram S. A New Kind of Science. – Wolfram Media, 2002.

ФИЛОСОФСКО-МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ РАЗВИТИЯ ИНТЕЛЛЕКТУАЛЬНОЙ РОБОТОТЕХНИКИ PHILOSOPHICAL AND METHODOLOGICAL PROBLEMS OF THE DEVELOPMENT OF INTELLIGENT ROBOTICS

泥其其拿 «智能机器人发展的哲学和方式
论问题»

Никитина Елена Александровна

*координатор молодежных проектов НСММИ РАН,
доктор философских наук, профессор кафедры
философии, социологии и политологии Московского
технологического университета (МИРЭА), г. Москва*

Nikitina Elena A.

*Doctor of Philosophy, Professor, Moscow Technological
University (MIREA)*



Аннотация. В статье систематизированы философско-методологические проблемы развития интеллектуальной робототехники. Проведен анализ проблем естественного интеллекта, социально-философских аспектов развития интеллектуальной робототехники, философско-психологических проблем взаимодействия человека и робота. Рассмотрены актуальные проблемы робоэтики. Обоснована необходимость социально-гуманитарной экспертизы научно-технических проектов с участием робототехнических систем.

Ключевые слова: эпистемология, информационно-коммуникационные технологии, когнитивные технологии, роботы, интеллектуальная робототехника.

Annotation. The article analyses the main philosophical and methodological problems of development of intelligent robotics. Author considers problems of natural intelligence, socio-philosophical aspects of development of intellectual robotics, philosophical and psychological problems of interaction between a man and a robot. Issues of Robo-ethics are considered. Author advocates the necessity of socio-humanitarian expertise of scientific-technical projects involving robotic systems.

Keywords: epistemology, intelligence, cognitive technology, robots, intelligent robotics.

Парадоксальным образом новейшие тенденции развития интеллектуальной робототехники возвращают нас к классической трактовке субъекта как бесстрастного, абсолютного инструмента получения объективной истины. Именно интеллектуальная робототехника в условиях техногенной цивилизации, интенсивной конвергенции высоких технологий [1] становится (на начальных этапах своего развития) технической реализацией модели классического субъекта познания. Интеллектуальный робот как «субъект» – это, разумеется, метафора. Тем не менее, в философском аспекте создание роботов с целенаправленным поведением, умеющих различать объекты внешней среды и воздействовать на них, обладающих определенными интеллектуальными способностями и управляющими системами, способных решать задачи общего характера, может быть представлено как сборка «субъекта» и техническая реализация определенных сторон человеческой деятельности.

Философия интеллектуальной робототехники как самостоятельное направление исследований оформилась в философии в последнее десятилетие [2, 3]. Выделим основные группы философских проблем интеллектуальной робототехники.

Философско-методологические проблемы естественного интеллекта. К данной группе относятся философские и методологические проблемы исследований естественного интеллекта человека, коллективного и социального интеллекта, проблемы соотношения субъектности и интеллекта, взаимодействия естественного и искусственного интеллекта в условиях распространения эргатических систем управления с гибридным интеллектом в различных видах человеческой деятельности. В группу эпистемологических проблем конвергентного развития когнитивных и информационных технологий входят вопросы взаимодействия эпистемологии и когнитивной науки, формирования трансдисциплинарной рациональности как нового типа рациональности [3, 4].

Философское осмысление естественного интеллекта в настоящее время определяется следующими факторами: 1) многочисленными экспериментальными данными о разнообразных видах и проявлениях интеллектуальной деятельности, накопленными в психологии, психофизиологии, когнитивной науке, нейронауках, искусственном интеллекте и других когнитивных дисциплинах; 2) необходимостью систематизации, обобщения и включения этих данных в корпус современных эпистемологических знаний; 3) отсутствием

общей теории интеллекта в психологии и потребностью в онтологии интеллекта, так как интеллект рассматривается в психологии преимущественно как процесс; 4) развитием когнитивной науки и когнитивных дисциплин, реализующих вычислительный подход к интеллекту; 5) диверсификацией современной эпистемологии под влиянием растущей дифференциации научного знания, что создало предпосылки для систематизации и интеграции экспериментальных данных об интеллекте в эпистемологию [3, 4].

Важно, что в настоящее время сложились предпосылки для интеграции представлений об интеллекте в различных науках. В психологии, в частности, достаточно широко распространено понимание естественного интеллекта как способности, объединяющей познание и целостную деятельность человека. Интеллект – это «... 1) общая способность к познанию и решению проблем, определяющая успешность любой деятельности и лежащая в основе др. способность; 2) система всех познавательных (когнитивных) способностей индивида: ощущения, восприятия, памяти, представления, мышления, воображения; 3) способность к решению проблем без проб и ошибок «в уме». Понятие И. как общая умственная способность применяется в качестве обобщения поведенческих характеристик, связанных с успешной адаптацией к новым жизненным задачам» [5].

В философии интеллект (лат. intellectus – ум, рассудок, разум) определяется как абстрактно-логическое мышление, как «...способность к опосредованному, абстрактному познанию, включающая в себя такие функции, как сравнение, абстрагирование, образование понятий, суждение, умозаключение; противостоит непосредственным видам познания – чувственному и интуитивному; в психологии – рациональное, подчиненное законам логики мышление; противостоит нерациональным сферам психики – эмоциям, воображению, воле и т.д.» [6].

Исследования в области искусственного интеллекта (ИИ) позволили выделить совокупность способностей, образующих естественный интеллект: «...(1) способность выделять существенное в наличных данных и знаниях и упорядочивать их...; (2) способность к целеполаганию и планированию поведения...; (3) способность к отбору знаний (посылок выводов, релевантных цели рассуждения); (4) способность извлекать следствия из имеющихся фактов и знаний...; (5) способность к аргументированному принятию решений, использующему упорядоченные знания (представление знаний) и результаты

рассуждений, соответствующие поставленной цели; (6) способность к рефлексии – оценке знаний и действий; (7) наличие познавательного любопытства: познающий субъект должен быть способен задавать вопрос «что такое?» и искать на него ответ; (8) способность и потребность находить объяснение (не обязательно дедуктивное!), как ответ на вопрос «почему?»; (9) способность к синтезу познавательных процедур, образующих эвристики решения задач и рассмотрения проблем...; (10) способность к обучению и использованию памяти; (11) способность к рационализации идей: стремление и умение уточнить их как понятия; (12) способность к созданию целостной картины относительно предмета мышления, объединяющей знания, релевантные поставленной цели (т.е. формирование приближенной «теории» предметной области); (13) способность к адаптации в условиях изменения жизненных ситуаций и знаний, что означает коррекцию «теории» и поведения» [7].

В философско-методологическом аспекте существенный интерес представляет изучение эволюции подходов к моделированию интеллекта в когнитивной науке. В методологии когнитивной науки проявилась тенденция смены классической когнитивистской вычислительной парадигмы динамическим, синергетическим подходом. Классический вычислительный подход был нацелен на получение точного знания о познавательных процессах, на создание гипотетико-дедуктивной модели, обеспечивающей возможность экспериментальной проверки (технической реализации) гипотез о познании. Данный методологический подход подвергся критике за сведение многообразия познавательных способностей человека к интеллектуальным, формализуемым функциям, за рассмотрение интеллекта вне связи с целеполагающей деятельностью человека и его телесностью. В модели познания, основанной на вычислительном подходе, интеллектуальные функции представлены как завершенные. Эволюционное происхождение данных функций, особенности их формирования в онтогенезе, социокультурные контексты познания и активность субъекта не учитывались, что снижало объяснительный потенциал данного подхода.

Безусловно, моделирование целостной, интеллектуальной, целеполагающей деятельности человека – это сложная задача, к реализации которой научное сообщество только приступает. На данном этапе исследований значимыми являются результаты моделирования познавательных способностей других живых существ. В частности,

интерес представляет роевой интеллект, применяемый в управлении роботизированными средствами. Роевой интеллект характерен для стайных действий живых существ – групп насекомых, птиц, рыб, обменивающихся информацией во время совместных действий по выполнению общей задачи. Предполагается, что со временем часть задач по управлению децентрализованной системой (рой беспилотных летательных аппаратов, группа наземных роботов) перейдет от оператора к роевому интеллекту, т.е. будет «запущена» самоорганизующаяся техническая система. Принцип роевого интеллекта применяется, в частности, для организации поисково-спасательных операций. Сервисные роботы-спасатели действуют с учетом данных, поступающих от других роботов, т.е. действие обеспечивается «распределенным» интеллектом. Выделяют также коллективные стратегии децентрализованного управления.

Целесообразно сопоставить эволюцию представлений о субъекте в философии и изменения в методологической платформе когнитивной науки. В современной эпистемологии субъект трактуется как конкретный телесный индивид, существующий в пространстве и времени, включенный в определенную культуру, имеющий биографию, находящийся в коммуникативных, социальных и иных отношениях с другими людьми. Субъект познания принадлежит природе и социуму, он деятелен и рефлексивен, отражает и конструирует мир, понимает и интерпретирует его, принимает решения, действует и достигает поставленных целей [8, 9].

Методология когнитивной науки в настоящее время эволюционирует в направлении интегративного видения человека и моделирования деятельности человека. Моделирование интеллекта ориентировано на учет деятельностной природы человека и социокультурных контекстов интеллектуальной деятельности, что способствует включению в когнитивную науку философской антропологии, герменевтики. В синергетическом искусственном интеллекте внимание привлечено к коллективным аспектам интеллектуальной деятельности – таковы гибридные интеллектуальные системы, системы «мягких вычислений», распределенные системы управления, интеллектуальные агенты, многоагентные системы, виртуальные организации, компьютерное моделирование общества на основе агентоориентированных моделей, эволюционирующие искусственные сообщества. Исследуются процессы формирования, деятельности, коммуникации, эволюции и кооперации искусственных систем [10, 11, 13, 14].

Происходящие методологические изменения дают дополнительные основания для утверждения о том, что развитие интеллектуальной робототехники является, в философском аспекте, своеобразной верификацией (технической реализацией) модели субъекта.

Для понимания направления и перспектив развития интеллектуальной робототехники важен принцип единства индивидуального, коллективного (микросоциального) и социального (макросоциального) субъектов, или, в другой формулировке, единства индивидуальных, коллективных и социальных когнитивных структур [9, 12]. В качестве эпистемологически значимых структур социальной реальности, формирующих субъектность и интеллект (объектов моделирования), могут рассматриваться социальные институты, стереотипы поведения, деятельности, нормы и т.п., которые осознанно/неосознанно осваиваются индивидуальным субъектом, становятся его личностными когнитивными установками и одновременно значимым фактором функционирования социума, влияют на содержание и форму знания. Многое усваивается индивидом без осознания, без рефлексии и таким же способом функционирует в виде сложившихся у индивида когнитивных функциональных систем и установок, которые могут реализоваться в деятельности индивида автоматически, под влиянием привычных «запросов» среды и потребностей человека. Существует единство сознания, бессознательного и деятельности. Важно также отметить, что универсальным основанием взаимопонимания в процессе коммуникации является диалогичность познания, обеспечивающая социальное взаимодействие [9, 12].

Компьютерная революция, широкое распространение информационно-коммуникационных технологий (ИКТ), развитие интеллектуальной робототехники в различных видах деятельности человека привели к качественно новому соотношению познавательной деятельности и информационных технологий, на основе которых формируется, воспроизводится и развивается познание человеком мира, создаются, хранятся, используются и передаются следующим поколениям знания. Наблюдается конвергенция когнитивных и информационных технологий.

Вместе с тем конвергентное развитие технологий проектируется, реализуется в междисциплинарном сообществе. Одной из характерных особенностей коммуникации в полидисциплинарных сообществах является конфликт интерпретаций, возникающий в связи с различием значений употребляемых терминов.

Так, в настоящее время широко используется термин «когнитивные технологии». В когнитивной науке, информатике, компьютерных науках под когнитивными технологиями понимаются интеллектуальные компьютерные технологии работы с большими массивами слабоструктурированной и неструктурированной информации (в частности, нейрокомпьютинг). В социально-гуманитарном знании термин «когнитивные технологии» употребляется, в большинстве случаев, в значении «технологии познания», в связи с возросшей рационализацией и «технологизацией» познания. Действительно, познавательная деятельность человека, использующего когнитивные компьютерные технологии, частично адаптируется к новым информационным технологиям, «технологизируется». Собственно, именно этот процесс в социально-гуманитарном знании и трактуется как когнитивные технологии в широком смысле, что и порождает конфликт интерпретаций в условиях междисциплинарного взаимодействия.

В эпистемологии под когнитивными технологиями предлагается понимать социокультурно обусловленные и исторически изменчивые формы адаптации познания к информационно-коммуникационным технологиям, меняющимся вместе с информационными революциями. Это одно из направлений эволюции естественного интеллекта. Соответственно, когнитивная культура личности – это эффективная технология преобразования информации в знание в соответствии с целями, смыслами и ценностями человеческой деятельности. Формирование когнитивной культуры является одной из существенных задач системы образования, ведь от сферы высшего образования общество должно получать не только профессионала, но и человека, обладающего навыками аналитического, критического, рефлексивного мышления, знающего методологию и умеющего применять методы, владеющего навыками коммуникации, диалога, взаимодействия с другими людьми на основе этических норм, умеющего принимать обоснованные решения, т.е. человека высокой интеллектуальной духовной культуры [9, 12].

Социально-философские и философско-психологические проблемы взаимодействия человека и робота. Данная группа взаимосвязанных проблем сформировалась и приобрела актуальность в связи с развитием социальной и бытовой робототехники. Социальными роботами называют роботов различного предназначения, включенных в повседневную жизнь человека и социальную сферу. Вместе с тем критерии отнесения роботов к социальным еще не окончательно

сформировались [15, 16]. Являются ли роботы, помогающие человеку решать бытовые проблемы, такие как уборка дома, приготовление пищи, помощь в сервировке стола, охрана дома (Roomba, Scooba, Mahru-Z и др.), социальными или сервисными?

Очевидно, что неотъемлемыми социальными функциями социального интеллектуального робота являются общение с человеком в процессе совместной деятельности, планирование поведения, управление целенаправленными действиями, обучаемость. Социально-коммуникативный робот должен обладать способностью к пониманию людей и себя в социальном контексте, должен уметь выражать и распознавать эмоции, устанавливать и поддерживать социальные отношения и социальные компетенции. Интерактивное (социальное) взаимодействие робота и человека должно строиться на основе социальных норм и правил. Социальный робот должен быть оснащен социокультурно ориентированным интерфейсом, соответствующим его функции и целевой аудитории.

Возникают вопросы: можно ли проектировать степени «социальности» робота? Можно ли и нужно ли задавать степень, меру «социальности» робота в связи с функциями, для выполнения которых робот разрабатывается, и его предполагаемым местом в социальной иерархии (робот-помощник, робот-компаньон (член команды), робот-руководитель, робот-учитель)? Как будут трансформироваться представления о социальном взаимодействии и социальной коммуникации в связи с ростом применения интеллектуальных систем и интеллектуальной робототехники в различных сферах жизнедеятельности общества?

Группа философско-психологических проблем связана с внешним видом и конструкцией механики роботов, используемых в социальной сфере: роботы-тележки, роботы-домашние животные, антропоморфные роботы (гуманоидные роботы). От решения данных проблем в немалой степени зависит социальная интеграция роботов и социально-психологическая адаптация человека к роботам, ведь эволюционно сформированная архаическая мифологичность нашего мышления создает яркие образы опасностей и угроз, исходящих от неживых существ, наделенных интеллектуальными способностями. Данные проблемы могут рассматриваться в антропологии техники.

Робозтика. Вопросы, связанные с морально-этическими и социальными аспектами внедрения роботов в жизнедеятельность

общества, рассматривались на Первом международном симпозиуме по робоэтике (Сан-Ремо, Италия, 2004 г.) и были отражены в принятом «Манифесте робоэтики». В чем состоит специфика роботов как моральных агентов? Должны ли разработчики роботов нести морально-этическую ответственность за «деятельность» роботов? Нужно ли наделять роботов способностью самостоятельно принимать решения? Как научить роботов уважать достоинство, свободу и права человека? Как обеспечить эффективную и безопасную коммуникацию людей и роботов?

В робоэтике активно обсуждается модель нравственности бытовых роботов, к которым относятся социальные роботы, роботы-няни, роботы-помощники и др., при этом предполагается, что специализация роботов, нацеленная на выполнение ограниченного круга задач, будет определять число норм, которые необходимо заложить в робота. Для определения таких норм будет использоваться опыт взаимодействия роботов, выполняющих определённые функции, с человеком. В частности, опыт работы роботов с пожилыми людьми показывает, что робот не должен быть эгоистичным, главное для него – интересы человека. Негативные эмоции, такие как зависть, гнев, страх, обида, не должны влиять на моральные решения робота, деформируя их. Роботы должны быть терпеливыми, не должны бездействовать. Робот, решая поставленные задачи, может и должен действовать в соответствии с принятыми в социуме нормами [13, 16, 17]. Вместе с тем, даже если удастся запрограммировать данные желаемые характеристики, смогут ли роботы-сиделки полноценно выполнять свою социальную функцию? Выслушают ли они пожилого человека, проявят ли интерес к его жизни, накопленной мудрости? Сумеют ли воспринять эту мудрость, сохранить в социальной памяти и передать следующим поколениям? Как научить социальных роботов моральному поведению и самостоятельности в принятии решений, как это делает человек в условиях изменчивой и вероятностной среды?

Данные задачи высокого уровня сложности в настоящее время трудновыполнимы. Более того, есть основания полагать, что развитие социальной робототехники может привести к утрате некоторых механизмов воспроизводства социальности.

Гуманитарная и социальная экспертиза научно-технических проектов, связанных с разработкой интеллектуальных робототехнических систем. Развитие интеллектуальных систем, интеллектуальной

робототехники, расширение областей их применения (военная сфера, здравоохранение, поисково-спасательные операции и т.д.), работы в направлении создания интеллектуальной адаптивной среды человеческого существования (Интернет вещей) сопровождаются обсуждением новых рисков человеческого существования, негативных и позитивных сценариев развития. Наиболее обсуждаемым комплексом вопросов, относящихся к гуманитарной и социальной экспертизе научно-технических проектов с участием интеллектуальной робототехники, являются вопросы применения автономных роботов в военной сфере. В частности, вопросы создания международно-правового комплекса и философско-этического кодекса, регламентирующих применение интеллектуальных роботов в военных технологиях. Допустимо ли применение боевых автономных роботов против людей в вооруженных конфликтах? Приемлема ли самооборона роботов в случае попытки их уничтожения противником? Каковы социальные и гуманитарные последствия интенсивного развития военной робототехники? Единого сформированного общественного мнения по данным вопросам не существует. Соответственно, актуальной является задача систематизации аргументов сторонников и противников военной робототехники и развитие общественного диалога всех заинтересованных сторон.

В заключение хотелось бы обратить внимание на растущую актуальность философско-методологических исследований проблем развития интеллектуальной робототехники и динамичность данного направления, обусловленную инновационными трансформациями современного общества.

ЛИТЕРАТУРА

1. Конвергенция биологических, информационных, нано- и когнитивных технологий: вызов философии (материалы круглого стола) – Вопросы философии. – 2012. – № 12. – с.3-23.
2. Всероссийская междисциплинарная конференция студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации». Сайт конференции. [Электронный ресурс]. URL: <http://www.scmaintconf.ru>
3. Алексеева И. Ю., Никитина Е. А. Интеллект и технологии. М.: Проспект, 2016. 96 с.
4. Никитина Е. А. Конвергентные технологии и трансформация структуры познания // Образовательные ресурсы и технологии. 2014. № 5 (8). С.157-166.
5. Большой психологический словарь. Под ред. Б.Г.Мещерякова и В. П. Зинченко. М.: АСТ, ПраймЕврознак, 2009. С. 816.

6. *Суворов О. В.* Интеллект // Новая философская энциклопедия: в 4 т. / Ин-т философии РАН; Нац. обществ.-науч. фонд; Предс. научно-ред. совета В. С. Степин. — М.: Мысль, 2000—2001.
7. *Финн В. К.* К структурной когнитологии: феноменология сознания с точки зрения искусственного интеллекта // Вопросы философии. 2009. № 1. С. 88-103.
8. *Никитина Е. А.* Проблема субъекта познания в современной эпистемологии // Перспективы науки и образования. 2015. № 2 (14). С. 16-24.
9. *Никитина Е. А.* Субъект познания, когнитивная культура личности и образование как Hi-Hume // Ценности и смыслы. 2011. № 7 (16). С. 94-108.
10. *Редько В. Г.* Моделирование когнитивной эволюции: на пути к теории эволюционного мышления. — М.: Ленанд, 2015. 256 с.
11. *Макаров В. Л.* Искусственные общества // Экономика и математические общества. 2012, Т. 48, № 3. С. 3-20.
12. *Никитина Е. А.* В поисках утраченного единства сознания: исчезло ли «Я»? Вестник Московского университета. Серия 7: Философия. 2009. № 5. С. 31-44.
13. Искусственный интеллект: философия, методология, инновации. Сборник трудов IX Всероссийской конференции студентов, аспирантов и молодых ученых. Москва, МИРЭА, 10-11 декабря 2015 г. Под общ. ред. Е. А. Никитиной. — М.: МИРЭА, 2015. 360 с.
14. *Лохин В. М., Манько С. В., Романов М. П., Диане С. А. К.* Перспективы применения, принципы построения и проблемы разработки мультиагентных робототехнических систем // XII Всероссийское совещание по проблемам управления. ВСПУ – 2014. 16-19 июня 2014 г. – Москва, с. 3810-3821.
15. *Зильберман Н. Н.* Функциональная классификация социальных роботов // Гуманитарная информатика. 2014. № 8. С. 30-39.
16. *Hegel F. et al.* Understanding social robots // Advances in Computer-Human Interactions. 2009. ACHI'09. Second International Conferences on. IEEE. 2009. P. 169–174.
17. Сайт «Робоцентр», Сколково [электронный ресурс]. URL: <http://sk.ru/foundation/itc/robotics/> (Дата обращения 14.11.16).

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ПСИХОЛОГИЧЕСКИЙ ВЫЗОВ

ARTIFICIAL INTELLIGENCE AS A PSYCHOLOGICAL CHALLENGE

业尔说法 «人工智能作为一个心理办法»

Ершова
Регина Вячеславовна

*доктор психологических наук, профессор,
Государственный социально-гуманитарный
университет, Коломна*

Ershova Regina V.

*Doctor of Psychology, Professor, State Social and
Humanitarian University, Kolomna*



Аннотация. В статье рассматриваются две группы вызовов, связанных с развитием современных цифровых технологий, в том числе связанных с искусственным интеллектом. Описываются актуальные последствия: социальные, социально-психологические, психологические, нейропсихологические «технологического прорыва», делается вывод о неоднозначности наблюдаемых изменений, необходимости глубоких исследований в данной области.

Ключевые слова: искусственный интеллект, общение, когнитивные нарушения, эмоциональные нарушения, академические достижения.

Annotation. In the article, two groups of challenges related to the development of modern digital technologies, including those related to artificial intelligence, are considered. The actual consequences are described: social, socio-psychological, psychological, neuropsychological «technological breakthrough», a conclusion is made about the ambiguity of the observed changes, the need for in-depth studies in this field.

Keywords: Artificial intelligence, communication, cognitive impairment, emotional disorders, academic achievements.

Приблизительно в 1950 году в обществе произошел так называемый информационный переход, или информационная революция, следствием чего стало глубокое проникновение информационных технологий во многие сферы человеческой деятельности, создание нового информационно-технологического контекста культурно-исторического развития человека [1, 2, 6, 7, 8, 9]. «Удивительно, как быстро сформировался совершенно новый тип среды, где вкус, обоняние и осязание не стимулируются, где большую часть времени мы сидим у экранов, а не гуляем на свежем воздухе и не проводим время в разговорах лицом к лицу», — пишет британский нейробиолог, профессор Оксфордского университета Сьюзен Гринфилд [13].

Новая среда, зачастую связываемая с искусственным интеллектом (точнее, цифровыми технологиями), задавая базовый вектор развития, ставит перед человеком и человечеством ряд вызовов, игнорировать которые становится все более и более затруднительно.

Первая группа вызовов – социальные и социально-психологические. Общение постепенно перемещается в виртуальное пространство без времени, без границ: с одной стороны, технологии во много раз упростили процесс информационного обмена. Начало этому процессу было положено профессором Нью-Йоркского университета Самюэлем Морзе, которому в 1837 году удалось впервые передать на расстояние электрический сигнал. Едва ли не первым, кто уже тогда смутно уловил специфику и последствия нового изобретения, был Генри Дейвид Торо, писатель и философ. Услышав о том, что с помощью телеграфа человек из штата Мэн сможет мгновенно переслать сообщение своему корреспонденту в Техасе, он задумчиво спросил: «И что же они скажут друг другу?». В этом вопросе предвосхищен анонимный характер наступавшей эпохи информационных технологий. Обогащение одной стороны привело к деформации, редукции, уплощению перцептивной и интерактивной сторон. В анонимном виртуальном пространстве происходит расщепление образа коммуникативного партнера, снижение социального интеллекта (очевидно, что интенсивное интеллектуальное и творческое развитие не гарантирует успешности социального поведения).

Изобретение ротационной печатной машины, фотоаппарата, телевидения в корне изменило характер передаваемой информации, сделав ее единицей не слово, а визуальный знак. Информация перестала быть пропозициональной и сделалась репрезентативной:

в отличие от высказывания, изображение нельзя ни подтвердить, ни опровергнуть – оно физически предъявляется зрителю и никак не ориентировано относительно истины. Иконическая информация, обращенная к максимально широкой, а значит анонимной аудитории, не требует селективной, логической и концептуализирующей работы разума, она апеллирует непосредственно к эмоциям и к способности сознания мгновенно распознавать паттерны. Телевидение заменило доказательное рассуждение, аргументированное изложение отвлеченной мысли, выработанные высокой книжной культурой, «картинкой» и «разговорами».

Новая символическая среда, вызванная к жизни «графической революцией», неумолимо разъедает социальные основы организации общества. Во-первых, для того чтобы усвоить визуальную (телевизионную) культуру во всей ее полноте, не требуется стадии обучения. Перед ее лицом семи- и семидесятисемилетний человек в принципе равны. В подобной визуальной культуре не существует понятия грамотности, ведь в визуальном «языке» нет алфавита. Во-вторых, эта культура не предъявляет сколько-нибудь серьезных и сложных требований к рациональной деятельности сознания и формам поведения человека. В-третьих, общество, которое само отождествляется с аудиторией масс-медиа, все быстрее теряет различительные культурные признаки и становится однородным. Так, в Америке появилась новая классификация возрастов, в которой традиционные стадии возрастного развития заменены всего тремя: младенчество, полувзрослый-полудетский возраст (*adult-child*), глубокая старость (*senility*) [15]. Напомним, что эволюционное преимущество сложности наглядно продемонстрировал Нобелевский лауреат, экономист Фридрих фон Хайек.

Высочайшее развитие технологий привело к резкому повышению продолжительности человеческой жизни: медицина спасает и поддерживает слабых и больных. Происходит экспоненциальное накопление генетического груза: каждое поколение биологически слабее предыдущего и все в большей степени зависит от искусственной среды. Медики всерьез говорят о накоплении «усталости генов». Наиболее отчетливо данная проблема проявляется на уровне детской популяции. По данным Д. И. Фельдштейна, в настоящее время происходит медиализация детей, выражающаяся в том, что детям ставят диагнозы, которые ранее ставили только взрослым. Каждые десять лет на 10-15% возрастает распространенность основных

форм психических заболеваний детей. Нервно-психические заболевания являются причиной 70 % инвалидности с детства, у 20% детей наблюдается минимальная мозговая дисфункция. В некоторых регионах страны количественно начинают преобладать дети, которых по нейропсихологическим показателям относят к группе «пограничной между нормой и патологией» (клинические, пограничные, субнормативные варианты развития), общее психическое недоразвитие (олигофрения) наблюдается у 22,5 % детей, дисгармоничное развитие (психопатии) у 26,5 % [8].

Вторая группа вызовов – вызовы индивидуального уровня, связанные с очевидными изменениями, которые происходят в логике и содержании развития личности. В 80-х годах прошлого века в научной литературе появился новый термин – цифровой человек (digital native) – именно так публицист Марк Пренски предложил называть детей, родившихся в цифровую эпоху и не представляющих своей жизни без цифровых средств массовой информации и коммуникации.

Информатизацию часто связывают с исключением, отмиранием ранее сформированных, но впоследствии ставших ненужными навыков, умений, видов и форм деятельности. Для обозначения указанного явления может быть использован термин «экзудия» (лат. exutio – исключение, изъятие, истребление). Примерами экзудии являются многочисленные данные о постепенном угасании вычислительных навыков (устного счета, складывания и умножения «в столбик», извлечения корней, возведения в степень и т.д.). Имеются разрозненные сведения о том, что информатизация, стимулируя «наиболее удобные» для формализации виды деятельности, способствует постепенному угасанию «конкурирующих» способов познания. Так, проведенные в 2007 году в Тринити колледже Дублина исследования показали, что 25 % его участников в возрасте до 30 лет не смогли вспомнить номера своих домашних телефонов, только 40 % помнят наизусть дни рождения близких родственников в сравнении с 87 % респондентов старше 50 лет. Ian Robertson назвал этот феномен «спровоцированной технологиями атрофией памяти» [9].

В многочисленных исследованиях доказано, что чрезмерное использование интернета детьми и подростками может привести к нарушениям внимания (ADHD), снижению эффективности когнитивного контроля [5]; у старших дошкольников, которые много времени проводят у экранов мониторов, обнаружены проблемы с освоением

чтения [13,14]. Turkle рассматривает анимизм как базовую особенность мышления интернет-зависимых людей [17].

Очевидным следствием выявленных трудностей является снижение академических достижений. Angrist и Lavi [11] отметили снижение успеваемости по математике у четвероклассников, по многим учебным предметам у старшеклассников после введения в школе компьютеров. Немецкие исследователи Т. Фукс и Л. Вёссман, сопоставив учебные достижения детей (полученные в ходе тестирования по программе PISA (Program for international student assessment) и наличие у них компьютера, пришли к выводу, что дети, у которых дома компьютер, хуже читают, считают, успевают в школе [3]. Исследования в США показали, что у использующих средства мультимедийной электронной коммуникации (ТВ, компьютер, видеоигры и др.) дошкольников нарушается внимание, а в старшем дошкольном возрасте возникают проблемы с чтением, у подростков усиливается социальная изоляция [9].

В 2007 году южнокорейские психологи обнаружили у детей и подростков Южной Кореи когнитивные и эмоциональные нарушения, которые были названы «цифровым слабоумием» (digital dementia). Цифровое слабоумие у подростков проявляется как потеря памяти, расстройство внимания, когнитивные нарушения, подавленность и депрессия, низкий уровень самоконтроля. Исследования показали, что в мозгу таких детей наблюдаются изменения, схожие с теми, что появляются после черепно-мозговой травмы или на ранней стадии деменции — слабоумия, которое обычно развивается в старческом возрасте [9].

Исследователи объясняют подобные нарушения изменениями в развитии головного мозга, являющимися следствием раннего взаимодействия с цифровыми технологиями и нового формата обучения, опирающегося на их использование. Так, Zhou Y. и др. обнаружили специфические изменения в структуре головного мозга (уменьшение серого вещества в левом полушарии: передней и задней частях поясной извилины, островка и медиальной затылочно-височной извилины) у интернет-зависимых подростков [18].

Еще одной проблемой развития информационных технологий является взаимопроникновение и взаимовлияние, слияние двух интеллектов: очеловечивание, наделение машины функциями сознания и омашинивание, технологизация сознания человеческого. Человек постепенно передает функции регуляции (саморегуляции)

вовне, машине. Многочисленная техника все более сокращает объем ручного, направленного на самообслуживание труда, технологичные гаджеты существенно расширяют возможности коммуникации и структурирования времени, медицинские препараты обеспечивают не только сохранение и поддержание здоровья, но становятся средством регулирования эмоциональной, сексуальной активности, циклов сна-бодрствования и др. Таким образом, человек «деятельный», активно преобразующий пространство вокруг себя, имеет большие шансы превратиться в человека «реагирующего», классически вписывающегося в бихевиористскую стимул-реактивную модель Дж. Уотсона. Наличие параллельной виртуальной среды обеспечивает бегство от проблем. Теперь уже не нужно столь усердно тренировать и развивать себя, достаточно начать играть в «Секонд лайф».

Активное использование современных технологий связывается с нарушениями в социальной и коммуникативной сферах: терапевты конституируют появление новых невротических расстройств — социальной тревоги (страха отключения от сети), социофобии, расширяется социальная изоляция, растут семейные разногласия, разводы, финансовые потери и утраты рабочего места [9, 12, 16]. Bushman и Anderson [10] обнаружили и описали обусловленный интернет-зависимостью поведенческий феномен (феномен мягкого насилия), проявляющийся в снижении эмпатии и сострадания.

Вместе с тем далеко не все прогнозы об отмирании различных форм и видов традиционной деятельности реализуются в действительности. Применительно к коммуникативной деятельности очевидно, что информатизация не только не приводит к сужению сферы общения, а, напротив, способствует развитию и расширению связей между людьми за счет расширения круга коммуникативных партнеров, развития опыта социальных контактов (деловых и личностных), нахождения новых оснований и причин для вступления в общение. Следует констатировать необходимость углубленного изучения позитивных сторон влияния современной технологической среды на психику человека.

В своих работах А. Назаретян формулирует закон техно-гуманитарного баланса, раскрывающийся как закономерная зависимость между тремя переменными: «силой», «мудростью» и «жизнеспособностью» общества. Очевидно, что мощные технологии требуют более совершенных форм культурной регуляции, в целях сохранения общества. Этот закон реализовывался на протяжении всей человеческой истории. С ним, на наш взгляд, связаны и надежды на нынешнее выживание человечества.

ЛИТЕРАТУРА

1. Бабаева Ю. Д., Войскунский А. Е. Психологические последствия информатизации // Психологический журнал, Т.19, №1, 1998, с.89-100
2. Вейценбаум Дж. Возможности вычислительных машин и человеческий разум // От суждений к вычислениям. М.: 1982.
3. Вессман Л., Ханучек Е. Роль качества образования в экономическом росте // Вопросы образования, 2007, № 2, С. 86-116.
4. Ершова Р. В. Психологические последствия технологического прорыва // в сб. Россия и мир: Развитие цивилизаций. – М.: ИМЦ. – 2014. С.113-124
5. Ершова Р. В., Варченко Н. Н., Ганькин К. А. О психофизиологических предикторах личностных свойств // Человеческий капитал. -№7 (67). 2014, С.52-55
6. Лейперт С. Переворот в сознании: дети, компьютеры и плодотворные идеи. М.: 1989.
7. Тихомиров О. К., Бабаева Ю. Д., Войскунский А. Е. Общение, опосредствованное компьютером // Вестник МГУ. Сер. 14. Психология. 1989. N 3.
8. Фельдштейн Д. И. Глубинные изменения современного Детства и обусловленная ими актуализация психолого-педагогических проблем развития образования // Вестник практической психологии образования. -№1, 2011, с.45-54.
9. Шнурцер М. Антимозг: цифровые технологии и мозг. – Litres, 2014.
10. Anderson C. A. et al. Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review // Psychological bulletin. – 2010. – Т. 136. – №. 2. – С. 151.
11. Angrist J., Lavy V., Schlosser A. Multiple experiments for the causal link between the quantity and quality of children // Journal of Labor Economics. – 2010. – Т. 28. – №. 4. – С. 773-824.
12. Brod C. Technostress. The human cost of the computer revolution. Addison-Wesley, 1984
13. Greenfield S. A. The human brain: A guided tour. – Basic Books, 2008.
14. Kolb B., Whishaw I. Q. Fundamentals of human neuropsychology. – Macmillan, 2009.
15. Neil Postman. The Disappearance of Childhood – New York: Vintage Books, a division of Random House, Inc., 1994.
16. Small G., Vorgan G. iBrain. – Harper Collins, 2009.
17. Turkle S. The second self: Computers and the human spirit. – Mit Press, 2005.
18. Yen J. Y. et al. The comorbid psychiatric symptoms of Internet addiction: attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility // Journal of adolescent health. – 2007. – Т. 41. – №. 1. – P. 93-98.

МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ МНОГОАГЕНТНЫХ СИСТЕМ

METHODOLOGICAL PROBLEMS OF MULTI-AGENT SYSTEMS

设寺它可瓦 《多代理系统的方式问题》

Шестакова

Марина Анатольевна

кандидат философских наук, доцент кафедры
философии и методологии науки Московского
государственного университета им.
М. В. Ломоносова, г. Москва

Shestakova Marina A.

PhD in Philosophical sciences, Associate professor,
Department of Philosophy and Methodology of
Science, Faculty of Philosophy of the Lomonosov
Moscow State University, Moscow



Аннотация. Изучение многоагентных систем (МАС) представляет собой широкую междисциплинарную область исследований. Одной из актуальных задач является включение отдельных аспектов конфликтологии в теорию и методологию МАС. Принцип ограниченной рациональности рассматривается как связующее звено между конфликтологией и теорией МАС. Ставится вопрос о перспективах исследования манипуляций и ресурсных конфликтов в рамках МАС.

Ключевые слова: искусственный интеллект, многоагентные системы, конфликтология.

Annotation. The study of multi-agent systems (MAS) is a broad interdisciplinary field of study. One of the important tasks is to incorporate certain aspects of conflictology into the theory and methodology of MAS. The principle of bounded rationality is regarded as a link between the conflictology and MAS. On perspectives of studies of manipulation and resource conflict are being placed within the problematic of MAS.

Keywords: multi-agent systems, artificial intelligence, conflictology.

Определяя приоритетные направления развития искусственного интеллекта (ИИ), Д.А.Поспелов наряду с другими «горячими точками» выделил многоагентные системы (МАС)[9]. В широком смысле под МАС понимается коллектив взаимодействующих агентов, совместные действия которых направлены на решение единой сложной задачи. Исследование МАС представляет собой междисциплинарную область, поскольку в качестве коллектива взаимодействующих агентов может рассматриваться биологическое сообщество (стая птиц, например), социальная или техническая система. Агентами МАС могут выступать как искусственные, так и естественные объекты: животные, люди, роботы, компьютерные программы. Междисциплинарный характер исследований МАС предполагает необходимость и возможность привлечения наработок других наук в теорию МАС. В настоящей статье рассматривается вопрос о включении отдельных аспектов конфликтологии в область моделирования МАС.

Одной из базовых категорий в теории МАС является понятие агента. Модель агента описывается через указание ряда признаков таких как, например, автономность, способность к коммуникации, наличие памяти и др. Существует большое количество различных моделей и классификаций агентов. Исходя из задач настоящей статьи, мы остановимся лишь на одной из них. По типу поведения агенты могут быть разделены на две группы: реактивные и интеллектуальные. Реактивные агенты действуют по схеме «стимул-реакция», отличаются слабой индивидуальностью и сильно зависят от окружающей среды. Интеллектуальные агенты часто наделяются целым рядом антропоморфных признаков, таких как социальное поведение, убеждения, желания, благожелательность, правдивость, лживость и др. В отличие от реактивных интеллектуальные агенты имеют «ярче выраженную индивидуальность и характеризуются целесообразным поведением в сообществе агентов, а также стремлением использовать ресурсы других агентов для достижения собственных целей»[13, с.106]. Приведенная классификация агентов позволяет выделить два основных направления в современных исследованиях МАС: распределенный искусственный интеллект (РИИ) и искусственную жизнь (ИЖ). МАС первого типа представляет собой объединение небольшого числа интеллектуальных агентов, каждый из которых выполняет свою задачу, сформулированную в результате декомпозиции (распределения) общей задачи (цели). Во втором случае речь идет о

совместной деятельности большого количества неинтеллектуальных (реактивных) или слабоинтеллектуальных агентов, объединившихся для достижения целей определенного типа, в основном для решения адаптивных задач «на выживание».

Стоит отметить, что в РИИ, также как и в ИЖ, агенты отличаются неполнотой информации о целях и задачах других агентов и системы в целом, а также ограниченным объемом ресурсов. Степень неполноты напрямую связана со степенью интеллектуальности агента. В предельном случае простой реактивный агент не обладает информацией о наличии «соседнего» агента, воспринимаемого им как часть окружающей среды. Другим пределом является ограниченность интеллектуального агента, который при любой степени интеллектуальности обладает лишь частью информации и других ресурсов, требуемых для решения общей задачи. Принципиальная ограниченность индивидуального агента не позволяет ему решить сложную задачу в одиночку, что с необходимостью ведет к объединению действий отдельных агентов в МАС. Неполнота информации агента часто называется ограниченной рациональностью. С нашей точки зрения, принцип ограниченной рациональности является связующим звеном между конфликтологией и теорией МАС.

Первоначально принцип ограниченной рациональности был сформулирован Г. Саймоном. Он поставил вопрос о том, что интеллект и внимание, необходимые человеку для решения сложных задач, представляют собой ограниченные ресурсы [10]. С подачи Г. Саймона понятие ограниченной рациональности широко используется в тех экономических теориях, которые рассматривают деятельность экономических агентов как последовательность актов принятия решений. В этих теориях ограниченная рациональность определяется как «характеристика экономических агентов, решающих задачу выбора в условиях неполной информации и ограниченных возможностей по ее переработке» [4, с.330]. Нужно отметить, что принцип ограниченной рациональности, как правило, используется вместе с принципом оппортунистического поведения, под которым имеется в виду действие агента, нацеленное на преследование собственного интереса и не ограниченное моральными нормами. Очевидно, что оба принципа, используемые в экономических теориях, *de facto* совпадают с двумя важнейшими характеристиками интеллектуальных агентов, входящих в модели МАС. Речь идет о том, что агент обладает

собственными целями, не совпадающими с целями других агентов, и принципиально ограничен в ресурсах.

Нужно отметить, что в экономической интерпретации принцип ограниченной рациональности связан с проблемой конфликта. Чтобы продемонстрировать это, можно еще раз обратиться к Г. Саймону. Он пишет о том, что современные экономисты больше не рассматривают экономику как науку о производстве, обмене и потреблении товаров и услуг. Сегодня для большинства экономистов более приемлемой оказывается трактовка экономики, предложенная в Международной энциклопедии социальных наук (International Encyclopedia of the Social Sciences), согласно которой «экономическая теория... изучает процесс распределения ограниченных ресурсов между множеством конкурирующих целей...»[10,с.16]. С нашей точки зрения, в вышеприведенном определении сама экономика представлена как способ разрешения или предупреждения ресурсных конфликтов. Таким образом, можно сказать, что принцип ограниченной рациональности подводит к проблеме конфликта и в экономической теории, и в области моделирования МАС. Различия в целях и ограниченность ресурсов интеллектуальных агентов МАС являются минимальными условиями для возникновения конфликтных ситуаций.

В современной конфликтологии используются различные определения конфликта. В самом общем виде конфликт определяется как «наиболее острый способ разрешения значимых противоречий, возникающих в процессе взаимодействия сторон»[1, с. 535]. Конфликт рассматривается как процесс, состоящий как минимум из трех стадий: предконфликтной, собственно конфликтной и стадии разрешения конфликта. Поэтому изучение минимальных условий конфликтов, предконфликтных ситуаций и потенциальных конфликтов входит в исследовательское поле конфликтологии. Здесь стоит обратить внимание на противостояние двух позиций в конфликтологии. Одна из них может быть названа узкой. Она рассматривает конфликты применительно к человеку и человеческому обществу. Другая – широкая трактовка, исследует конфликт как особый тип взаимоотношений, возникающий не только в человеческом обществе, но и в живой природе [1]. Есть и еще более общий взгляд на конфликт как на противоречие, имеющее место в том числе и в физическом мире [5, с.424]. Среди вариаций второго, широкого подхода нужно отметить так называемую единую теорию конфликта, цель которой заключается в разработке универсальной модели

конфликта и его математического описания [11]. В теории МАС востребована широкая трактовка конфликта, применимая к социальным, биологическим и техническим системам, к моделям РИИ и ИЖ.

В моделировании МАС, прежде всего, обращается внимание на создание условий, предотвращающих возникновение конфликтов. Так, например, интеллектуальные агенты могут наделяться желанием избегать конфликтов в качестве одного из своих свойств. Вместе с тем рассматриваются и сами конфликты. В литературе выделяют несколько типов конфликтов в МАС. Первый – это конфликты убеждений, возникающие в том случае, если агент получает от другого агента ложную информацию или противоречащую его убеждениям. Другой тип – конфликты рефлексии, возникающие из-за неполноты информации об окружающей среде и других агентах. Кроме того, выделяют конфликты целей и ресурсные конфликты [2]. Конфликты присущи как РИИ, так и ИЖ, но при этом часть конфликтов, например, конфликты убеждений, встречаются только в РИИ. Несмотря на то, что в МАС выделяется несколько типов конфликтов, реально исследуется в основном один – конфликт убеждений.

«При этом под конфликтом, обычно, понимают ситуацию, в которой возникает противоречие вида $p \dot{\cup} q \Rightarrow \text{false}$, где p и q – убеждения агентов». А под разрешением конфликта понимается «снятие логического противоречия вида $p \dot{\cup} q \Rightarrow \text{false}$ за счет отбрасывания одной из альтернатив в соответствии с некоторым критерием, или смены p и q вместе»[2].

Перспективы разработки темы конфликтов в теории и методологии МАС связаны, на наш взгляд, с развитием двух направлений. Первое из них касается формального описания конфликтных ситуаций. Здесь, как нам кажется, нужно обратить внимание исследование рефлексивных конфликтов, предложенное В. А. Лефевром. По существующим оценкам работа В. А. Лефевра «Конфликтующие структуры» «дает достаточно адекватные способы описания конфликтов данного вида, однако в настоящее время в известных работах эти модели вообще не принимаются во внимание» [2].

В настоящее время, по нашему мнению, остается актуальным замечание В. А. Лефевра о том, что утилитарные задачи по оптимальному решению конфликтов заслонили собой теоретическое исследование конфликта как особого феномена [7, с.2]. Теория конфликта предполагает разработку формальных способов описания конфликтных ситуаций различного типа. В. А. Лефевр еще в 60-е годы предложил

рефлексивный подход к изучению конфликтов. Он отталкивался от традиционного понимания рефлексии как способности человека встать в позицию наблюдателя по отношению к своему телу, действию и мышлению. И придал понятию рефлексии универсальный смысл. «В самом общем виде рефлексия в понимании Лефевра – это способность некоторых систем строить модели себя и одновременно видеть себя строящими такие модели» [6,с.29]. Предложенный подход позволил выделить рефлексия в качестве обособленного предмета исследования и формально (алгебраически) описать рефлексивные процессы, независимо от свойств носителей рефлексии.

Понятие рефлексии было использовано В.А.Лефевром в концепции рефлексивного управления. Это иной взгляд на управление, отличающийся от традиционного кибернетического подхода. При рефлексивном подходе объект управления наделяется рефлексией, т.е. имеет некоторое представление о себе самом, о другом объекте, входящем в систему. В модели рефлексивного управления отношения между субъектом и объектом управления трансформируются в субъект-субъектные. Существенным свойством такого управления является то, что оно не предполагает обязательного наличия обратной связи. В.А.Лефевр дает различные определения рефлексивного управления. Под ним имеется в виду «процесс передачи оснований для принятия решения одним из персонажей другому..»[7, с. 25], «воздействие на субъектов, склоняющее их принять решения, заранее подготовленные управляющей стороной» [12,с.54]. Стоит согласиться с мнением о том, что у концепции рефлексивного управления есть твердое ядро – манипулятивное управление [12]. Для нашей темы важно подчеркнуть связь между конфликтом, манипуляцией и рефлексивным управлением.

Эту связь можно проследить, сравнив подходы В.А.Лефевра и Т. Шеллинга [14]. Последний предложил создать модель конфликта, используя теорию игр. Он обратил внимание на то, что стратегические игры предполагают такую взаимозависимость решений противников, когда наилучший образ действий игрока зависит от действий другого игрока. Этот принцип Т. Шеллинг использовал в своей концепции конфликта, подчеркивая, что в ходе конфликта можно управлять поведением других или оказывать на него воздействие. Свой подход он иллюстрирует, в частности, на примере стратегии сдерживания. Смысл сдерживания он видит в том, чтобы «оказать влияние на выбор, делаемый другой стороной, что достигается путем

оказания влияния на ее ожидания относительно того, как мы будем себя вести. Сдерживание также включает в себя предъявление противнику свидетельств, заставляющих его поверить в то, что наше поведение будет определяться его поведением» [14, с. 26]. На наш взгляд, концепции Т. Шеллинга и В. А. Лефевра близки друг другу. И в том, и в другом случае речь идет о манипулировании противником, о том, чтобы заставить его «свободно» выбирать «правильную» линию поведения. Подобная стратегия позволяет удерживать конфликт в латентном состоянии, не доводя его до открытого столкновения. С другой стороны, нужно подчеркнуть, что концепция рефлексивного управления В. А. Лефевра имеет специфические отличия. В ней делается акцент на рефлексивной стороне взаимодействия противников. «Участники конфликта рассматриваются как игроки, вступившие в своеобразную рефлексивную игру. Термин «рефлексивный» подчеркивает, что игроки отражают в мышлении рассуждения друг друга» [8, с. 8]. Преимуществом концепции рефлексивного управления является формализация, что позволяет ее использовать в теории МАС. Моделирование конфликтов манипулятивного типа можно, как нам кажется, отнести к актуальным задачам исследования МАС. В этом контексте остается востребованной модель рефлексивного управления В. А. Лефевра.

Второй вопрос, связанный с разработкой темы конфликтов в теории МАС, касается «физической» стороны конфликтов. В литературе по проблемам искусственного интеллекта давно заявлено о «телесной» составляющей ИИ. Х. Дрейфус заметил, что в исследованиях по моделированию интеллекта не учитывается «роль тела в организации и интеграции нашего опыта» [3, с. 334]. Он справедливо отмечал, что разумное поведение человека не сводится к вычислительным операциям, мозг связан с телом человека, поэтому в моделировании интеллектуальных функций необходимо учитывать телесную составляющую. Х. Дрейфус исходит из общей способности человека приобретать телесные умения и навыки. Благодаря этой способности мы можем, например, научиться использовать различные орудия труда и инструменты. Это обучение не предполагает обязательного знания и применения правил логики или других рациональных норм. Оно может протекать как приспособление тела к орудиям труда. В этой связи Х. Дрейфус правомерно ставит вопрос о существовании «информационных процессов», основанных на телесной организации человека. Например, освоение родного языка происходит не через

заучивание правил, а через «вживание» в язык. В информационных процессах этого типа значение целого обладает приоритетом по отношению к составляющим его элементам. Х. Дрейфус приходит к выводу о том, что телесная организация человека позволяет ему выполнять такие разумные функции, для которых пока нет машинных программ. Иными словами, работа по созданию ИИ должна проводиться с учетом того, что тело является необходимым условием существования человеческого интеллекта. Вопрос о «телесной» стороне интеллектуальной деятельности нужно, как нам кажется, поставить и применительно к моделированию МАС. Одним из способов учесть «физическую» составляющую может, с нашей точки зрения, стать разработка модели ресурсного конфликта, возникающего между интеллектуальными агентами МАС.

В моделировании МАС распределение ресурсов между агентами обычно подчинено достижению общей цели, решению сложной задачи. В соответствии с этим рассчитывается объем средств, необходимых агенту для выполнения поставленной перед ним локальной подзадачи. Логика такого подхода понятна, но она слишком проста для того, чтобы описывать взаимоотношения между высокоинтеллектуальными агентами. Такого рода агенты составляют, например, социальные МАС. Они могут накапливать ресурсы «про запас», не тратя их на решение общих задач, что может служить основанием для ресурсного конфликта с другими агентами. К примеру, затраты на содержание управляющей организации хозяйствующего субъекта могут многократно превышать уровень, необходимый для выполнения этой организацией своих функциональных обязанностей. Это пример того, что в социальной системе может образовываться нечто вроде «жировых отложений», ресурсная диспропорция, которая продиктована скорее целями «выживания» отдельных агентов, чем решением задач, поставленных перед системой в целом. Снятие конфликтов, возникших на почве ресурсов, видимо, не может быть сведено к устранению формально-логических противоречий в убеждениях агентов. Здесь стоит заметить, что и В.А. Лефевр, о котором шла речь выше, обратил внимание на физическую сторону конфликта. «Конфликт мыслящих, – пишет он, – во многих случаях проявляется через «физическое» взаимодействие. Поэтому помимо логических средств исследования интеллектуальной деятельности необходимо строить теоретические модели различных «физических» взаимодействий конфликтующих систем» [7, с.2]. Как нам кажется, в этой

функции может выступить модель ресурсного конфликта, которая позволяет учитывать нерациональные, «телесные», «физические» цели интеллектуальных агентов в моделях МАС.

Подводя итог, отметим, что современные исследования МАС опираются на целый ряд концепций: теорию деятельности, теорию коммуникации, теорию игр и др. В этом ряду есть место и для конфликтологии. Тема конфликта была выделена и Д. А. Поспеловым. Говоря о горячих точках в исследовании ИИ, он обратил внимание на возможное создание нового научного направления – «теории асинхронных конфликтующих процессов или что-то подобное с другим названием, которое еще не появилось» [9].

ЛИТЕРАТУРА

1. Аницупов А. Я., Шипилов А. И. Конфликтология: Учебник для вузов. – М.: ЮНИТИ, 2000.
2. Городецкий В. И., Грушинский М. С., Хабалов А. В. Многоагентные системы (обзор)/ <http://spkurdyumov.ru/networks/mnogoagentnye-sistemy-obzor/> дата обращения 12.11.2016.
3. Дрейфус Х. Чего не могут вычислительные машины.- М.: Прогресс, 1978.
4. Институциональная экономика. Новая институциональная экономическая теория. Учебник под ред. А. Аузана. – М.: ИНФРА-М, 2006.
5. Конфликтология: учебник для студентов вузов, обучающихся по специальностям экономики и управления и гуманитарным специальностям. Под ред. В. П. Ратникова. – М.: Юнити-Дана, 2012.
6. Лепский В. Е. Лефевр и рефлексия/Рефлексивные процессы и управление. Международный научно-практический междисциплинарный журнал № 1, 2006, январь-июнь, том 6, с.26-37.
7. Лефевр В. А. Конфликтующие структуры. Издание второе, переработанное и дополненное. – М.: Изд-во «Советское радио», 1973.
8. Лефевр В. А., Смолян Г. Л. Алгебра конфликта, М.: Знание, 1968.
9. Поспелов Д. А. Десять «горячих точек» в исследованиях по искусственному интеллекту/Интеллектуальные системы (МГУ). – 1996. – Т.1, вып.1-4. – С.47-56.
10. Саймон Г. Рациональность как процесс и продукт мышления/THESIS, 1993, вып. 3, с.16-38.
11. Светлов В., Семенов В. Конфликтология: Учебное пособие.- СПб.: Питер, 2011.
12. Смолян Г. Л. Рефлексивное управление – технология принятия манипулятивных решений/Труды ИСА РАН. Том 63, 2/2013, с.54-61.
13. Тарасов В. Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика.- М.:Эдиториал УРСС, 2002.
14. Шеллинг Т. Стратегия конфликта. М.: ИРИСЭН, 2007.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ НАУКИ: ПЕРСПЕКТИВА АНТИРЕПРЕЗЕНТАЦИОНАЛИЗМА

ARTIFICIAL INTELLIGENCE AND COGNITIVE SCIENCE: ANTIREPRESENTATIONALIST PERSPECTIVE

米骇落伏 «人工智能与认知科学：反展示
性的远景»

*Михайлов
Игорь Феликсович*

*кандидат философских наук, научный
сотрудник Института философии РАН,
г. Москва*

Mikhailov, Igor F.

*CS in Philosophy., Senior Researcher, Institute of
Philosophy RAS, Moscow*



Аннотация. В статье прослеживаются параллели в историческом и проблемном развитии теории искусственного интеллекта и когнитивных наук. Вместе с тем делается попытка обосновать антирепрезентационализм как наиболее перспективную позицию в обоих научных направлениях.

Ключевые слова: искусственный интеллект, когнитивные науки, коннекционизм, социальные сети, репрезентационализм, антирепрезентационализм.

Annotation. The article traces the parallels in historical and problematic development of the theory of artificial intelligence and cognitive science. At the same time, the author strives to justify antirepresentationalism as the most promising position in both scientific fields.

Keywords: artificial intelligence, cognitive science, connectionism, social networks, representationalism, antirepresentationalism.

Когнитивные науки и теории искусственного интеллекта объединяет сущностная взаимосвязь. Во-первых, теоретически ИИ обычно рассматривается как составная часть комплекса когнитивных наук. И во-вторых, оба дисциплинарные направления стали возможны благодаря бурному развитию информационно-вычислительных технологий в середине XX в.

Развитие когнитивных наук началось в середине XX-го века в качестве ответа на засилье бихевиоризма в психологии. Бихевиоризм формировался в рамках общей позитивистской тенденции, когда не было научных средств и возможностей изучать внутреннее содержание мозга и сознания. В этих условиях точная наука могла ориентироваться только на изучение наблюдаемого поведения. В 50–60-е годы XX в. появляются первые компьютерные технологии. Однако задолго до этого, ещё в XVII веке, Томас Гоббс провозгласил девиз: «Рассуждения есть расчёт». В 1830-е гг. Чарльз Бэббидж высказал идею универсальной автоматической вычислительной машины. В 1936 г. Алан Тьюринг писал, что открываются возможности изобретения одной [универсальной] машины, которую можно использовать для вычисления любой вычислимой последовательности [8].

В 1976 г. Ньюэлл и Саймон выдвинули гипотезу «физической символической системы» как машины, которая развивает во времени некий набор символических структур. «Такая машина (будь то человек или цифровой компьютер) имеет необходимые и достаточные условия для общей разумной деятельности» [3, 116]. Один из классиков символической парадигмы в когнитивной науке Зенон Пилишин писал в 1984 г.: «Люди способны... действовать на основе репрезентаций, потому что они физически формируют такие репрезентации как когнитивные коды, и потому что их поведение является каузальным следствием операций, произведённых на основе этих кодов. Поскольку это именно то, что делают компьютеры, отсюда следует, что познание является разновидностью вычисления» [5, xiii].

Однако в истории параллельно развивался другой подход к структурированию материальных объектов. Этот подход я бы назвал сетевой парадигмой. В 1736 г. Леонард Эйлер решил задачу о семи кенигсбергских мостах. В результате родилась математическая теория графов, которая до сих пор является основным математическим инструментом для анализа сетевых структур.

В середине 1980-х гг. появляется коннекционизм как альтернативное направление в когнитивной науке. Он возник как

междисциплинарный подход к изучению познания, который интегрирует элементы из области искусственного интеллекта, нейрофизиологии, когнитивной психологии и философии сознания. В рамках этого подхода предполагается, что когнитивные явления можно объяснить с помощью набора общих принципов обработки информации, известной как параллельные распределенные вычисления (PDP) [6].

С методологической точки зрения, коннекционизм — это рабочая рамка для изучения когнитивных явлений с использованием архитектуры простых процессоров, соединенных между собой с помощью взвешенных соединений. Согласно этой модели, каждый нейрон получает множество входящих сигналов от других нейронов. Нейрон интегрирует сигналы путем вычисления взвешенной суммы активации. На основании величины полной входной активации функция активации (например, пороговая функция) определяет уровень исходящей активации нейрона. Исходящая активация распространяется на последующие нейроны.

Для того чтобы сеть была обучаемой, был внедрен алгоритм обратного распространения ошибки (backpropagation of error). Этот алгоритм может быть использован для контролируемого обучения многослойной сети. Для этого типа обучения алгоритм представляет сеть, которая попарно связывает входящие паттерны возбуждения и желаемые исходящие паттерны. Алгоритм вычисляет ошибку на выходе, то есть разность между фактическим выводом сети и желаемых паттернов. Далее алгоритм обратно распространяет соответствующие сигналы ошибки через каждый слой сети. Эти сигналы ошибки используются для определения изменений веса, необходимых для достижения минимизации ошибки вывода.

Группой исследователей, заложивших основы PDP, был поставлен эксперимент по обучению коннекционистской сети использованию неправильных английских глаголов в прошедшем времени. Лингвистам известен эффект так называемого U-turn в освоении людьми прошедшего времени неправильных глаголов. В процессе обучения коннекционистская сеть воспроизвела тот же самый эффект: то есть в начале эксперимента сеть обнаружила хорошее знание тех форм, которые принимают неправильные английские глаголы в прошедшем времени. Зачем качество этого понимания ухудшилось. И только на последней стадии обучения оно вернулось на прежний уровень. По мнению сторонников коннекционизма, этот эксперимент показывает, что их подход в когнитивной науке

указывает правильный путь, достаточно реалистично воспроизводя психологию реальных людей.

Вместе с тем методы исследования сетей находили себе дорогу и в социальной науке. В 1973 г. Марк Грановеттер опубликовал знаменитую статью «Сила слабых связей» [1]. В этой статье он пытался найти теоретический мостик между социологическими теориями низшего и высшего уровня. Таким связующим звеном, по мысли автора, должна стать теория, согласно которой социальные воздействия лучше всего распространяются и приводят к более существенным результатам, будучи транслируемы по так называемым слабым сетевым связям, в отличие от сильных — например, семейных — связей. Он писал: «В этой статье я буду доказывать, что анализ процессов в межличностных сетях составляет наиболее плодотворный мостик между микро- и макроуровнем. Тем или иным способом, именно через эти сети малые взаимодействия транслируются в более масштабные паттерны, которые, в свою очередь, влияют на малые группы» [1, 1360]. Далее он пишет: «Что бы ни распространялось, это может достичь большего количества людей и преодолеть большие социальные дистанции... если это распространяется через слабые связи, скорее чем через сильные» [1, 1366]. Интересно в этой связи то, что Грановеттер для построения теории использовал методологию и математический аппарат, применяемый для анализа сетевых связей, заложив, по мнению многих, основы социологии социальных сетей.

В качестве параллельного процесса можно рассмотреть историю создания искусственных нейронных сетей. В наше время устройства на основе искусственных нейронных сетей играют большую роль в распознавании образов и речи, а также в робототехнике. Работа в этом направлении началась с того, что в 1958 г. Фрэнк Розенблатт создал машину под названием «Перцептрон». Эта машина своей архитектурой имитировала человеческий мозг. Марвин Минский и Сеймур Пейперт в 1969 г. доказали ограниченность этого устройства. И только в 1980-е годы, с прогрессом технологий, начался ренессанс искусственных нейронных сетей, что было связано с задачами обороны, а также с ростом интереса к экспертным системам. С появлением нанотехнологий ожидается дальнейший прогресс в этой области.

Нельзя не заметить, что сеть нейронов мозга и социальная сеть структурно аналогичны: обе состоят из элементов, умеющих выполнять несложные функции и взвешивать связи с близлежащими элементами. Обе имеют когнитивные надстройки, которые возникают

эмерджентно: в случае с мозгом это так называемый когнитом — термин К. В. Анохина, в случае с социальной структурой — это когнитивные социальные сети. Если мы имеем машины одной и той же нейронной архитектуры, управляемые разными программами, то между ними возможен интерфейс — библиотека функций, переводящих команды одного языка в команды другого. Таким интерфейсом между мозгом и обществом выступает естественный язык, развитый людьми в процессе социальной эволюции.

То, что мы в нашей культуре называем мышлением, на самом деле, по мысли Л. С. Выготского, представляет собой внутреннюю речь — свёрнутое, интериоризированное движение смыслов, подчиняющееся явным и неявным правилам языка. Минимальные условия осмысленности языка, как и любой знаковой системы, сформулированы в аристотелевской логике. Это линейное, по своей архитектуре, мышление создает картину, отличную от того, как на самом деле работает мозг, и как на самом деле устроено общество. Именно здесь корень многих труднорешаемых проблем эпистемологии и социальной философии.

Хороший, с фактологической точки зрения, обзор основных концепций ИИ представлен в [2]. Автор отталкивается от наиболее широко распространённой классификации направлений или подходов в исследованиях: символического, коннекционистского и «воплощённого» (situated). Однако он предлагает более детальную классификацию, в которой выделяет такие группы подходов, как (1) символический, (2) основанный на «грубой силе», (3) основанный на знаниях, (4) прецедентный, (5) креативный, (6) биокомпьютационный, (7) динамический и (8) «воплощённый».

Первая группа (1) содержит разновидности традиционного символизма, известного также в литературе как классицизм или компьютеризм. Этот подход автор иллюстрирует следующей теоретической притчей: если ваша жизнь в селе Большие Скучищи безрадостна, и вы хотите переехать в город Веселовск, то, с точки зрения классического когнитивиста-символиста, это значит, что внутри вашего когнитивного аппарата есть физически идентифицируемое состояние, соответствующее восприятию указанного села как скучного, другое состояние, соответствующее восприятию указанного города как весёлого, третье состояние, соответствующее желанию переехать, и четвёртое, соответствующее поведенческому намерению.

Я бы от себя добавил, что каждая из этих репрезентаций должна, в свою очередь, распадаться на логические составные части: восприятие вот этого как села Большие Скучищи, восприятие чего бы то ни было как безрадостного, синтез этих двух восприятий и т.п. В пределе эта когнитивная «онтология» должна совпасть с грамматикой языка — откуда, видимо, происходит ярко выраженный интерес классических когнитивистов к грамматике и лингвистике. В таком подходе нет ничего невозможного, но главная проблема состоит в том, что на уровне нейрофизиологии пока не найдено никаких взаимно однозначных соответствий ментальных состояний предполагаемым «репрезентациям» в виде ансамблей возбуждённых нейронов. То есть у данной теории нет солидного эмпирического базиса в классическом понимании слова, но есть как раз то, что в принципе связывает когнитивную науку с ИИ — успехи в компьютерном моделировании когнитивных функций. И это обстоятельство ставит интересный эпистемологический вопрос: можно ли в принципе заменить фактическое подтверждение моделированием? Любимый пример функционалистов: братья Райт не могли отправить аэроплан в полёт, пока они заставляли его махать крыльями подобно птице. Только используя другой аэродинамический механизм, они подняли его в воздух. Согласно общепринятой интерпретации, этот пример говорит о том, что функция не зависит от материальной реализации. Но с другой стороны, он показывает и то, что, реализовав эту функцию, может быть, проще и эффективнее, мы ничего не добавим к нашему знанию о том пути, по которому в действительности шла эволюция. То есть к нашему знанию о том, как на самом деле летают птицы, и как на самом деле мыслят люди.

И еще одно соображение, даже более важное для теории ИИ: по-другому реализовав функцию полёта, братья Райт не смогли придать самолёту ту степень маневренности, которой обладают птицы. Иными словами, они реализовали не совсем ту же самую функцию.

Подход со стороны того, что автор называет «грубой силой» (2), ярче всего воплощён в шахматных компьютерах типа Deep Blue. Согласно его анализу, программы, соревнующиеся с шахматистами начиная с 1950-х годов, построены по принципу простого поиска, т.е. быстрого перебора возможных сценариев выхода из данного положения, в сочетании с функцией, отсекающей неблагоприятные варианты. Под «грубой силой» подразумевается ставка на решительное

повышение вычислительной мощности, которая в рамках данного подхода оказывается главной гарантией «интеллектуальности».

В тексте выдвигается возражение против такого понимания природы когнитивности, которое мне представляется важным, но не решающим: согласно данным психологов, новички в шахматах и мастера игры видят доску по-разному. Первые — как ансамбль отдельных фигур, вторые — как сочетания кластеров, на которые распадаются возможные решения [2, 209]. Соответственно, типичная компьютерная программа, играющая в шахматы, по когнитивной стратегии ближе к новичку и, если и выигрывает у гроссмейстера, то только за счёт скорости и объёма вычислений.

Мне это возражение представляется верным, но не принципиальным: программу, абстрактно говоря, всегда можно изменить в сторону большего когнитивного реализма. Важнее другое: игра в шахматы плохо подходит в качестве модели взаимодействия с миром, поскольку число возможных ходов в каждом конкретном положении игры всегда конечно, и, следовательно, понятно, что должно считаться завершением каждого поискового цикла. В реальной жизни число доступных опций может быть конечным, но трудно определимым, а может быть и бесконечным. И в таких случаях программу, ориентированную на простой перебор доступных вариантов, не спасёт никакая вычислительная мощность.

Подход, основанный на знаниях (3), равно как и прецедентный (4), стремится к большему психологическому реализму. Исходя из того, что интеллектуальные процессы человека погружены в некий массив изначальных знаний о мире, сторонники этой теории ИИ исходят из того, что мыслящую машину нужно сначала «кормить знаниями с ложечки», пока она не научится добывать их сама [2, 211]. Примерно так же должна поступать машина и согласно прецедентному подходу: она должна всякую новую ситуацию «видеть как» старую. В качестве технологических решений предлагаются «концептуальная зависимость» — подкладка концептуально-образной схемы мира под интеллектуальные процедуры — и «сценарии» (scripts) — наборы неких типичных операций по обработке информации, запускаемых при получении данных определённого рода.

Общая особенность этих подходов — стремление к достижению большего психологического реализма за счёт большей нагрузки на программиста. Мне трудно судить, насколько это всё оправдано с инженерной точки зрения, но как проект исследования человеческой

психики — вспомним аналогию с самолётами — эти подходы вряд ли на правильном пути.

Попытки моделировать творческую природу человеческого интеллекта (5) я не буду здесь обсуждать подробно, поскольку их методика, как она описана в обсуждаемой статье, сводится к действиям по аналогии и перекомбинированию известных паттернов — последнее используется в машинной имитации музыкальных стилей великих композиторов.

Под биокомпьютационным подходом (6) автор подразумевает, прежде всего, коннекционизм и, затем, «генетические алгоритмы». Описание последних осталось неясным для меня, поэтому они останутся здесь без обсуждения. О коннекционизме автор справедливо говорит, что «распределённый характер репрезентаций в коннекционистских моделях допускает обучение и появление нового поведения» [2, 219]. И это при том, что коннекционистским сетям нужен не программист, а, скорее, тренер. Казалось бы, этот подход преодолевает недостатки конкурирующих, но и здесь не обходится без порции критики. Оппоненты обвиняют коннекционизм в «сетевом мистицизме», «манящей непроницаемости эмерджентных систем» и даже в подверженности общему «постмодернистскому духу нашего времени» [2, 219]. Основа для этих обвинений — действительное несоответствие коннекционистских моделей требованиям алгоритмической разрешимости и трактативности, которые оппоненты считают общенаучными. Однако стоило бы возразить, что сами эти требования исходят из определённого понимания языка и логики как линейных правилосообразных систем. Коннекционистская архитектура исходит из принципиально иного понимания ментальных процессов — как сложных каузальных взаимодействий внутри нелинейных систем, осуществляющихся параллельно. И корректное теоретическое описание их предполагает использование скорее сложных математических моделей, чем логических исчислений и языков программирования.

Другой аргумент против коннекционизма состоит в том, что *de facto* в рамках этого подхода желаемый психологический реализм так и не достигается. Так, например, коннекционистские сети овладевают языковыми навыками до того, как демонстрируют «понимание» языка, тогда как у людей это происходит в обратном порядке. Но для того, чтобы принять или отклонить этот аргумент, необходимо ответить на сложный философский вопрос о природе понимания, а это

как раз одна из целей когнитивной науки, различные направления которой идут к ней различными путями. Витгенштейн, как известно, утверждал: мы можем говорить, что ребёнок понял, что такое натуральный ряд чисел, когда он научился продолжать его с любой позиции [9, sect. 143].

Больше ничего рационального и эмпирически верифицируемого о понимании сказать нельзя. Но это именно то, чего вполне реально можно добиться от коннекционистской сети: например, через обучение методом проб и ошибок она становится способной складывать звуки в осмысленные слова и образовывать грамматические формы. Точно так же, через глубинное обучение, можно научить её употреблять правильные слова в правильных контекстах — и витгенштейновский критерий понимания будет соблюден. Если мы не требуем большего от ребёнка для признания его понимающим, почему мы должны требовать большего от машины?

Динамический (7) подход и подход в рамках воплощённого (8) интеллекта я также объединил бы в одно семейство. Первый предполагает описание когнитивной деятельности в терминах «пространства» состояний системы, определяемых переменными, зависимыми от времени. Следствием такого подхода становится замена теоретического языка, связанного с вычислениями и репрезентациями, языком, связанным с геометрией и динамическими состояниями. Обобщая, можно было бы сказать, что в динамических объяснениях возрастает роль математики за счёт роли логики.

«Воплощённый» (situated) подход отказывается от классицистского понимания интеллекта как абстрактного, индивидуального, рационального и оторванного от восприятия и действия, противопоставляя этому понимание его как отелесненного (embodied), встроенного (embedded) и распределённого (distributed). Иными словами, когнитивные процессы протекают не в мозге, а между мозгом, остальным телом и средой. Классицистские подходы в ИИ пытаются создать искусственного эксперта или искусственного физика, тогда как начинать следовало бы с уровня «интеллигентности» насекомых, с тем чтобы на основе этих работающих моделей перейти к воссозданию языка и абстрактного мышления. Научиться завязывать шнурки машине труднее, чем научиться решать математические задачи или играть в шахматы, а ведь эволюции понадобились миллиарды лет для появления чувствительности и мобильности и только миллионы лет для выработки собственно человеческих способностей.

Я объединил последние два подхода в одну группу, поскольку, во-первых, у меня нет существенных возражений против их основных позиций. Во-вторых, я не нахожу в их принципах никаких противоречий с основами коннекционизма, что предполагает принципиальную совместимость всех трёх подходов. Их роднит:

- 1) принципиально антиклассицистская позиция — отказ от метафоры вычислений, возвращение к доброй старой каузальности;
- 2) поворот от логики к математике;
- 3) признание существенной роли биологической определённости человека в формировании его интеллектуальных качеств.

На мой взгляд, синтез этих трёх подходов имеет шанс составить наиболее перспективное направление в когнитивной науке и ИИ. Однако влияние динамического подхода на коннекционизм должно, на мой взгляд, привести последний к отказу от понятия репрезентации.

Обозначим ещё раз: классицистские подходы постулируют наличие принципиально идентифицируемых физических состояний когнитивной системы, взаимно однозначно соответствующих идентифицируемым ментальным состоянием. Эти репрезентации и представляют собой внутренние «символы», которыми оперирует сознание. Я же склонен считать, что вся сложнейшая машинерия мозга — как электрическая, так и химическая, — периферийной нервной системы, мышц и конечностей, социальных взаимодействий и коммуникации призвана обеспечивать сложное нелинейное поведение живых систем высокого порядка, в том числе языкового, благодаря которому и рождаются когнитивные или интенциональные концепты — «знать», «бояться», «видеть» и т.п. Попытка обнаружить непосредственные материальные референты этих терминов — например, в виде «репрезентаций» — есть признак наивного реализма.

В статье [4] предлагаются сильные и взаимосвязанные аргументы против репрезентационализма: эпистемологический и методологический. Первый состоит в том, что, даже если эволюция создала механизм, основанный на репрезентациях, не понятно, как мы могли бы его идентифицировать качестве такового. Второй утверждает, что само понятие репрезентации в когнитивной науке недостаточно операционализировано для того, чтобы можно было однозначно установить наличие или отсутствие репрезентаций в некоторой системе. На мой взгляд, отказ от репрезентационализма в когнитивной науке и ИИ мог бы опираться на следующую методологическую

позицию: паттерн возбуждения нейронов понимается как функция от данных ввода, а поведенческие команды — как функция от паттерна возбуждения. Ментальные состояния в таком случае сопоставляются не репрезентациям, а функциям. Тогда всё, что нужно от удовлетворительной научной теории сознания, — это математические формализмы (или модели), реалистично — в соответствии с фактами — описывающие обе функции. Реализация предложенного подхода означала бы окончательный «развод» языка науки с языком «народной психологии».

ЛИТЕРАТУРА

1. *Granovetter, Mark S.* The Strength of Weak Ties. In: *American Journal of Sociology*, 1973, 78(6):1360-1380.
2. *Ekbja, Hamid R.* Fifty years of research in artificial intelligence. In: *Annual Review of Information Science and Technology*, Volume 44, Issue 1, January 2010. Pages 201-242.
3. *Newell, A., & Simon, H.A.* Computer science as empirical inquiry. In: *Communications of the ACM*, 1976, 19(3), 113–126.
4. *Haselager, P., De Groot, A., Van Rappard, H.* Representationalism vs. anti-representationalism: a debate for the sake of appearance. In: *Philosophical Psychology*, Vol. 16, No. 1, 2003.
5. *Pylyshyn, Z.W.* Computation and cognition: Toward a foundation for cognitive science. Cambridge, MA: MIT Press, 1984.
6. *Rumelhart D.E., McClelland J.L., & the PDP Research Group.* Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume 1: Foundations. Cambridge, MA: MIT Press, 1986.
7. *Horgan, T.* Connectionism and the Philosophical Foundations of Cognitive Science. In: *Metaphilosophy*, January 1997, Volume 28, Issue 1-2. P. 13.
8. *Turing, A.M.* On computable numbers, with an application to the Entscheidungsproblem. In: *Proceedings of the London Mathematical Society*, 1936, 2(42), 230–265.
9. *Wittgenstein, L.* *Philosophical Investigations*. Oxford; Blackwell, 2009, 4th edition.

ГЕТЕРАРХИЧЕСКАЯ МОДЕЛЬ В РАМКАХ КОННЕКЦИОНИЗМА THE HETERARCHIC MODEL WITHIN THE FRAMEWORK OF CONNECTIVITY

养可扶丝卡压 «教育中的知识系统»

Янковская

Екатерина Алексеевна

*координатор образовательных проектов
НСМИИ РАН, кандидат философских наук,
ведущий научный сотрудник Российского
экономического университета имени
Г. В. Плеханова, г. Москва*

Yankovskaya Ekaterina A.

*Coordinator of educational projects of the RAS
SCMAI, candidate of philosophical sciences,
leading researcher of the Plekhanov Russian
Economic University, Moscow*



Аннотация. В статье рассматривается гетерархическая модель организации сетевых систем и потенциал ее применения для анализа структуры и принципов функционирования когнитивной системы. Показано, каким образом гетерархическая модель возникает в рамках коннекционизма благодаря работам одного из его основателей У. Маккаллока. Рассматриваются основные принципы гетерархической модели и возможность ее применения для исследования когнитивной системы.

Ключевые слова: Коннекционизм, гетерархия, когнитивная система, гетерогенность, динамичность, когнитивные процессы.

Annotation. The article considers the heterarchic model of organization of network systems and its potential application for the analysis of the structure and principles of the functioning of cognitive system. It is shown how the heterarchic model emerges within the framework of connectionism due to the work of one of its founders W. McCulloch. The main principles of the heterarchic model and the possibility of its application for the study of cognitive system are considered.

Keywords: Connectionism, heterarchy, cognitive system, heterogeneity, dynamism, cognitive processes.

Основные аспекты коннекционистской парадигмы

Коннекционизм ассоциируется, прежде всего, с разработками в области искусственного интеллекта, которые выступили в качестве альтернативы модельно-символическому подходу, реализованному на базе стандартной компьютерной архитектуры. Тем не менее, коннекционизм, как и модельно-символический подход, формирует целую парадигму в когнитивистике, в рамках которой разрабатывались теоретические и прикладные модели архитектуры когнитивных систем.

Когнитивная система в данном случае понимается в широком смысле слова как совокупность сложноорганизованных и взаимосвязанных когнитивных процессов (например, восприятия, памяти, мышления) и определенным образом (ре)презентированного ментального содержания, являющегося результатом действия этих процессов. Предполагается, что все эти составляющие когнитивной системы определенным образом структурированы, эта структура поддается формализации и может быть представлена в виде определенной когнитивной архитектуры. Коннекционизм предлагает определенные модели когнитивной системы, которые могут быть реализованы, в том числе, в искусственной среде.

Коннекционистский подход реализуется в построении искусственных нейронных сетей, предполагающих как написание особого программного обеспечения и использование специально разработанного математического аппарата, так и создание микросхем, имитирующих нейронные сети аппаратным путем. В настоящее время такой подход к созданию интеллектуальных систем снова стал популярен после того, как в течение достаточно долгого времени превалировал модельно-символический подход.

В основе коннекционистского подхода лежит стремление проанализировать и воспроизвести ключевые принципы устройства и функционирования нервной системы. Предполагается, что исследование того, как работает мозг, способно подсказать эффективные методы решения сложной задачи по моделированию когнитивной системы.

При построении искусственных нейронных сетей используется несколько основных свойств нервной системы, ставших базовыми принципами классического коннекционизма. Ключевое из этих свойств связано со способностью нейрона генерировать и передавать электрический импульс. Генерирование импульса осуществляется по

следующей схеме: если сумма всех входных импульсов превышает определенный порог, то нейрон создает собственный импульс, если нет – этого не происходит. Данное свойство нейрона становится источником представлений о том, что базовый уровень человеческого мышления очень прост: «Так или иначе, на низшем уровне нашего разума царит простое сложение. Перефразируя знаменитое изречение Декарта, «я мыслю, значит я суммирую» (от латинского *Cogito, ergo sum*) [1, с.321]. Таким образом, одним из основных свойств нервной системы считается функционирование ее элементарной составляющей – нейрона – по принципу «все или ничего», описываемого бинарной логикой.

Второе ключевое свойство нервной системы, используемое коннекционизмом, состоит в том, что исходные простые элементы – нейроны – объединены в сложную сеть. В этой сети возникает взаимное влияние элементов друг на друга, то есть стимулирование или торможение импульсов на различных участках сети. Получается, что за счет объединения простых элементов в сеть происходят сложные процессы.

Наконец, третий ключевой принцип коннекционизма состоит в том, что нейронные сети обладают пластичностью, они способны переструктурироваться в процессе обучения. За счет этого свойства искусственные нейронные сети не могут быть запрограммированы, как обычная компьютерная система. Правила функционирования системы формируются в ходе ее обучения на примерах: «в отличие от классических систем, которые используют явные, часто логические, правила, организованные в иерархии, для управления последовательной обработкой информации, представленной в виде символов, коннекционистские структуры опираются на параллельную обработку символических подсистем и используют статистические закономерности вместо логических правил преобразования» [5].

Итак, коннекционизм предлагает модель сложной обучающейся системы, базируясь на трех исходных принципах: основа системы – простые элементы (*units*), которые могут находиться в одном из возможных состояний, эти элементы объединены в сеть, которая обладает способностью переструктурироваться. Это абстрактная модель для описания естественной или искусственной нейронной сети как материального субстрата когнитивных функций. В то же время, коннекционизм предполагает, что свойства этого субстрата определяют

устройство когнитивной системы, то есть структуру когнитивных процессов и способы репрезентации их содержания.

Развитие коннекционизма позволило создать множество моделей функционирования искусственных нейронных сетей. В то же время, развитие нейрофизиологических исследований привело к тому, что были уточнены многие аспекты строения и функционирования мозга, которые оказались значительно более сложными, чем редуцированные коннекционистские модели. Тем не менее, в рамках коннекционизма стало возможно появление и развитие моделей когнитивных систем со сложной архитектурой. Примером такой модели можно считать гетерархическую модель, предложенную одним из основателей коннекционизма и кибернетики У. Маккаллоком.

Понятие «гетерархия» в концепции У. Маккаллока

Используя коннекционистское представление о том, что нейроны являются простыми элементами, принимающими одно из двух состояний, У. Маккаллоком построил систему формализации мышления, которая опиралась на булевы функции. Соответственно, формализованные нейроны могли выполнять цифровые функции, то есть воспроизводить формальную логику компьютера. В этом аспекте его коннекционистская концепция имела много общего с модельно-символическим подходом, представляющим мозг в виде компьютера особого типа.

Однако У. Маккаллоком обратил внимание и на особые свойства нервных сетей, которые не позволяют создать алгоритм работы нервной системы и выражаются в гетерархическом принципе ее действия. Возникновение понятие «гетерархия» связано с работой У. Маккаллока 1943 г. «Гетерархия ценностей, обусловленная топологией нервных сетей». В этой работе У. Маккаллоком указал на нетранзитивность когнитивных процессов. Нетранзитивность выражается в так называемой оценочной аномалии: «если организм обладает сложной нервной системой, ... то его поведение не может быть предсказано на основании строгой оценочной системы. Ему присуща гетерархическая оценочная система, которая слишком насыщена, чтобы подчиняться принципу *summum bonum*».[4, p.118] Оценочная аномалия описывается следующим образом: «А более предпочтительно, чем В, В более предпочтительно, чем С, но С более предпочтительно, чем А» ».[4, p.119]. Согласно У. Маккаллоку, такой парадоксальный способ выбора является основой устойчивого человеческого поведения.

Таким образом, когнитивная система, согласно У. Маккаллоку, не-транзитивна и рекурсивно замкнута. Разрабатывая данный принцип, Х. фон Ферстер указывал, что кругообразность является не только самоповторяющейся, но и продуктивной: «Мое намерение состоит в том, чтобы порочный круг *circulus vitiosus*... возвести на почетное место некоего *circulus creativus* – «созидающей кругообразности» [2, с.194]. Стабильная активность, согласно Х. фон Ферстеру, возможно только в самоперестраивающихся и самовоспроизводящихся системах.

Два исключаящие друг друга решения в данном случае постоянно сменяют друг друга по замкнутому циклу, иными словами, возможно сочетание сразу нескольких альтернатив, выбор одной из которых не предполагает полного отказа от остальных. Х. фон Ферстер указывает на то, что «... Его [Маккаллока] в нейрологическом аспекте интересовал вопрос о том, почему люди устроены таким образом, что оценочные аномалии принимаются ими в качестве так называемого нормального поведения. Он показал, что нервная система посредством «рефлекторной дуги» образует сенсомоторную петлю, замкнутый цикл из видения и действия, действия и видения, что в конечном итоге приводит к стабильной активности» [2, с.213]. Таким образом, основным качеством любого сложного процесса, связанного с когнитивными функциями, становится постоянное динамическое самовоспроизведение, самокоординация и самоизменение.

Для описания циркулярности гетерархии У. Маккаллоком предпочитал использовать не латинский термин «*circulus vitiosus*», а греческое понятие «*diallel*». Это связано с тем, что, в отличие от латинского понятия, несущего негативный оттенок, «термин «*diallel*» происходит от греческого понятия, *diallelos tropos*, он описывает вид кругообразности в контексте процесса ...и имеет динамический характер ... соответственно, бытие описывается через антиномичные процессы, а не состояния»[6]. Кругообразность здесь не предполагает «снятия» антиномичности в единстве синтеза, а состоит в постоянной динамической отсылке разнородных элементов друг к другу.

Взаимная координация различных уровней системы возможна за счет того, что У. Маккаллоком называл «избыточностью командных пунктов», то есть достаточной «компетентностью» и сложностью различных уровней системы. Избыточность гетерархии коррелирует с избыточностью нервной системы, которые были обнаружены в процессе нейрофизиологических исследований.

Идеи У. Маккаллока относительно гетерархической структуры когнитивной архитектуры достаточно долгое время не получали развития в рамках когнитивистики. Тем не менее, понятие «гетерархия» периодически появлялось в теории менеджмента, где использовались кибернетические идеи применительно к социальным системам, в частности, к функционированию фирм. В биологических исследованиях также периодически использовалось данное понятие. За счет развития понятия «гетерархия» в этих областях его можно снова применять в качестве модели когнитивной системы. Это возможно, поскольку при исследовании различных сложных систем можно обнаружить общие принципы их функционирования. Так, например, кибернетик С. Бир причислял экономику, социальные системы, когнитивный аппарат и живые организмы к типу очень сложных вероятностных систем.

Основные принципы гетерархической модели

Этимология понятия «гетерархия» отсылает к греческой терминологии и представляет собой соединение слов «έτερος» («другой», «чужой») и «ἀρχή» («начало», «власть», «управление»). Термин «гетерархия» буквально обозначает «многоначалие», «то, что находится под управлением другого», «распределенное управление» (в противоположность «священному порядку» иерархии») или же, в более свободном переводе, «нечто, обладающее внутренней разнородностью». Таким образом, это понятие является контингентным и отчасти альтернативным по отношению к понятию «иерархия». Гетерархия – это сетевая структура, однако она не является простой сетью равнозначных элементов, которые находятся в хаотическом состоянии. Это структура, где существует определенный порядок и правила взаимодействия между элементами, но эти правила не являются предзаданными и ригидными, как в иерархических системах.

Существует два основных принципа, которым подчиняются системы гетерархического типа: гетерогенность при сохранении общей целостности и динамичность при сохранении устойчивости. Гетерогенность гетерархической системы обеспечивается децентрализованной или временно централизованной структурой, рекурсивностью, гибкостью элементов и уровней. Динамичность гетерархической системы обеспечивается итеративностью, балансом положительных и отрицательных обратных связей, вероятностным характером функционирования системы.

Гетерогенная целостность гетерархической системы поддерживается за счет системы взаимосвязей между ее уровнями и элементами, которые взаимно влияют друг на друга. В гетерархической системе осуществляется не односторонняя, от высшего уровня к низшему, а взаимная, внутренняя координация действий, когда «высший уровень воздействует на низший уровень, будучи в то же время сам определен этим низшим уровнем». Сложность каждого уровня поддерживается всей сложностью системы, и, напротив, – сложность системы рекурсивно создается за счет сложности уровней. Нельзя определить исходный простой элемент или базовый управляющий уровень гетерархии, каждый элемент сам достаточно сложен, динамичен и задается сложностью всей системы.

Для поддержания рекурсивной продуктивности и взаимозависимости между уровнями нужна особая динамичность всей системы, поскольку единственный способ разрешения противоречия внутри гетерархии состоит в том, чтобы рассматривать ситуацию взаимоисключающего выбора в динамической форме. Таким образом, основным качеством любого сложного процесса, связанного с когнитивными функциями, становится постоянное динамическое самовоспроизведение, самокоординация и самоизменение.

Таким образом, иерархия и гетерархия не представляют собой совершенных противоположностей: системы обоих типов противостоят хаосу. Гетерархическая система, как и иерархическая, стремится к организованной активности. В то же время устойчивость гетерархической системы основана не на статичной стабильности (*stability*) как абсолютной неизменности, а на устойчивости к ошибкам (*robustness*). Если стабильность предполагает ригидность системы, которая не склонна изменять свое устройство в изменившихся условиях, то устойчивость к ошибкам возникает как раз за счет способности системы перестраиваться в изменившемся контексте, то есть реагировать на него. Устойчивость к ошибкам предполагает, что система не допустит фатального сбоя даже в том случае, если некоторое влияние среды будет достаточно дестабилизирующим: «Гетерархия является динамической системой, способной к улучшению со временем и обладающей устойчивостью к ошибкам» [3, p.23].

Применение гетерархической модели к анализу когнитивной системы

В настоящий момент не создана концепция, которая бы полностью описывала структуру когнитивной системы с помощью гетерархической модели. К тому же, сама данная модель нуждается в уточнении, детализации и формализации. Тем не менее, можно продемонстрировать несколько общих аспектов устройства когнитивной системы, если рассматривать ее как гетерархически структурированную.

Прежде всего, целостность и гетерогенность когнитивной системы предполагает, что когнитивные процессы не редуцируются друг к другу. Соответственно, нецелесообразно рассматривать «высшие» когнитивные процессы (связанные с функционированием неокортекса) в отрыве от более «примитивных» структур. Уровни иерархии «высших» и «низших» форм когнитивной деятельности являются спутанными, они рекурсивно взаимодействуют и определяют друг друга, причем это взаимодействие идет в различных направлениях. Также те процессы, которые исходно не относились к познавательным, например, эмоции, в ходе функциональной перестройки уровней системы способны выполнять когнитивные функции.

Информация, получаемая обоими полушариями, одна и та же, но обрабатывается она по-разному. Функции полушарий не только разделяются, но и перекрещиваются, что выглядит как избыточность. Мозг функционирует при этом как единое целое, интегрируя оба типа мышления как две взаимодополняющие подсистемы обработки когнитивной информации. Для каждого полушария характерен свой тип частичного автоматизма, и они не могут полностью контролироваться друг другом.

Также считается, что филогенетически «первичное» мышление людей – это мышление преимущественно образное, то есть правополушарное. В то же время дифференциация функций мозга обеспечила новые возможности и для эволюции образного мышления. Развитие знакового и логико-вербального мышления изменяет и образное мировоззрение, но, в свою очередь, не существует без невербальных форм. Все это говорит о том, что жесткой модулярности когнитивных процессов, по всей видимости, не существует.

В данном случае приведен только один пример того, каким образом применение гетерархической модели способствует систематизации принципов структурирования и функционирования когнитивной системы. Для более дифференцированной модели когнитивной

системы необходима дальнейшая работа над разработкой и применением гетерархической модели.

ЛИТЕРАТУРА

1. *Хофштадтер Д.* Гёдель, Эшер, Бах: эта бесконечная гирлянда. Самара, 2001.
2. *Цоколов С. А.* Дискурс радикального конструктивизма. Мюнхен, 2000.
3. *Gunji Y.-P., Sasai K., Wakisaka S.* Abstract heterarchy: Time/state-scale re-entrant form. // *BioSystems*. – 2008 – №91
4. *McCulloch W. A.* Heterarchy of Values Determined by the Topology of Nervous Nets // *Bulletin of Mathematical Biophysics*. – 1943. – N 5.
5. *Medler D. A.* A Brief History of Connectionism // [электронный ресурс] http://neuron-ai.tuke.sk/NCS/VOL1/P3_html/vol1_3.html
6. *Von Goldammer E., Paul J., Newbury J.* Heterarchy – Hierarchy. Two complementary categories of description // [Электронный ресурс] (http://www.vordenker.de/heterarchy/a_heterarchy-e.pdf).

О СООТНОШЕНИИ ИСКУССТВЕННОГО И ЕСТЕСТВЕННОГО ИНТЕЛЛЕКТА

ON THE RELATION BETWEEN NATURAL AND ARTIFICIAL INTELLIGENCE

卡特类其可 «人工智能与大脑智能的相比»

Катречко Сергей Леонидович

кандидат философских наук, доцент Школы
философии, Факультет гуманитарных наук,
Национальный исследовательский университет
«Высшая школа экономики» (НИУ ВШЭ), г. Москва

Katrechko Sergey L.

Associate Professor, School of Philosophy, Faculty of
Humanities, National Research University Higher School
of Economics (HSE), Moscow



Аннотация. Статья посвящена проблеме концептуального соотношения человеческого ума (интеллекта) и искусственного интеллекта. Обсуждаются три принципиальных отличия человеческого ума от существующих систем ИИ: 1) математическое мышление человека (как «познание посредством конструирования понятий» (Кант); 2) проблема ментальных образов; 3) феномен сознания (декартовское *cogito*, кантовская трансцендентальная апперцепция).

Ключевые слова: кантовская концепция математики, проблема ментальных образов, голографический принцип сознания, Кант, Гуссерль, Гибсон, Хамфри.

Annotation. My paper is devoted to the problem of conceptual relation between the human mind (intellect) and artificial intelligence. In this regard in article three fundamental differences of human mind from the existing systems of AI are discussed: 1) the mathematical human thinking (Kant); 2) the problem of mental images; 3) the consciousness phenomenon.

Keywords: Kant's theory of mathematics, mental image problem, holographic principle (consciousness), Kant, Husserl, Gibson, Humphrey.

Последние достижения XXI в. в области искусственного интеллекта (ИИ): программы интеллектуального поиска (Google), распознавания текстов (FineReader) и речи, машинного перевода (Promt), игровые компьютерные программы, уже обыгрывающие сильнейших игроков (чемпионов мира и Европы) в шахматы и игру Го, и др., — свидетельствуют о том, что проблема искусственного интеллекта в том виде, как она была поставлена в середине XX в., в принципе решена. Конечно, это не полноценный естественный интеллект, имеющий универсальный характер своего применения, но принципиальных трудностей по созданию квази-универсального ИИ как объединения воедино существующих «частных» программ (в частности, перечисленных выше) не видно. Трудности здесь носят скорее технический характер: необходимость совместимости разных алгоритмов, недостаток мощности вычислений, памяти, энергообеспечения, etc.

Вместе с тем XXI в. распрощался с иллюзией [полного] подобия искусственного и естественного интеллекта: машинные программы не моделируют структуру и механизмы деятельности человеческого разума, а «моделируют» лишь конечные результаты его работы (например, решение задач), они подобны естественному интеллекту лишь функционально, т.е. добиваются сходных результатов, возможно, принципиально другими методами. Машинная программа переводит текст (речь) с одного языка на другой или играет на равных с человеком в шахматы, но это не означает, что она делает это точно так же, как это делает человек, т.е. «копирует» нейро-когнитивно-психологические процессы его мышления (mind). Здесь можно вспомнить известный мысленный эксперимент «китайская комната» Дж. Серля: для машинного перевода с китайского языка программа ИИ не обязательно должна понимать [как это делает человек] содержание (смысл) текста, она это может делать чисто синтаксически или с привлечением формальной семантики, отличающейся от человеческой понимающей «семантики». В этом смысле ложной оказывается базовая предпосылка серлевского вопрошания: от программ ИИ не следует требовать полного «семантического» моделирования человеческого разума, программа ИИ может и не понимать, что и как она переводит, но при этом вполне успешно осуществлять собственно перевод, например, с английского на русский. Программа автоматического доказательства теорем может вполне успешно (быстро) и формально правильно доказывать логику-математические

теоремы, реализуя алгоритмы полного перебора (например, алгоритм Британского музея), психологически чуждые для человеческого интеллекта, который рассуждает, скорее, в системах натурального вывода, хотя, заметим, что подобные системы и методы работы в логических исчислениях — лишь относительно недавно созданные техники мышления и их вряд ли можно считать естественными для человека¹.

Тем самым, в серлевских терминах, можно говорить о создании/построении в XXI в. слабого, а не сильного искусственного интеллекта как полноценной модели разума человека и, возможно, о снятии самой задачи построения сильного ИИ как неправильно поставленной, хотя и здесь к XXI в. достигнут определенный прогресс². При этом, как мы уже отмечали чуть выше, сейчас, конечно, не приходится говорить о создании полноценного (универсального) искусственного разума, который мог бы решать любую возникающую проблему. Существующие системы ИИ моделируют лишь отдельные интеллектуальные функции, хотя и делают это весьма успешно.

Это означает, что поставленная Тьюрингом проблема соизмеримости естественного и искусственного интеллекта (тест Тьюринга) в принципе решена, и поэтому мы от вопроса «Возможен ли ИИ и как он возможен?» переходим к вопрошанию о сходстве/различии человеческого разума и ИИ.

В самом общем виде (схематично) соотношение этих двух интеллектов таково. Искусственный интеллект — это лишь чистый интеллект, или рассудок (в кантовской терминологии), а разум (= естественный интеллект) человека более сложно структурирован и может быть задан следующей триадой познавательных способностей «чувство — рассудок — разум». Таким образом, несмотря на функциональную схожесть собственно интеллектуальных функций, естественный интеллект человека обладает рядом способностей (функций), отсутствующих в настоящее время у ИИ. Их первый класс связан с нижележащей для интеллекта сферой пред-рассудка, которая в классической философской традиции соотносится со срединной между чувством и

¹ Заметим, что предложенные примерно в это же время (середина XX в.) секвенциальные исчисления (Г. Генцен) и семантические таблицы (Э. Бэт) имеют промежуточный статус: способам их использования можно «научить» как человека, так и машину, что, однако, не означает, что они являются методами сугубо естественного интеллекта, которые моделируются машинным разумом.

² Речь идет прежде всего о самообучающихся комплексах на основе нейронных сетей.

рассудком способностью воображения, или образным мышлением (образным модусом сознания), частным (наиболее значимым) случаем которой выступает математика, представляющая собой (по Канту) синтез воображения и рассудка. Во-вторых, это более высокая сфера разума, в которой выделим (здесь) такие сверх-рассудочные способности, как сознание (apperцепции) и рефлексия. Соответственно, для создания «среднего/сильного» ИИ следует научиться моделировать и эти функции человеческого разума (сознания).

При этом можно указать по крайней мере на три принципиальных отличия человеческого ума (естественного интеллекта) от существующих систем ИИ: 1) математическое мышление человека как созерцательно-дискурсивную деятельность по конструированию понятий (Кант); 2) проблема ментальных образов (наличие ментальных образов в человеческом уме и отсутствие (необязательность) таковых в системах ИИ); 3) феномен сознания (декартовское *cogito*, кантовская трансцендентальная апперцепция), без которого нельзя помыслить человеческий ум, но которое отсутствует (или является необязательным компонентом) для существующих систем ИИ.

* * *

1. Начнем с анализа математической деятельности, опираясь на ее кантовское понимание как конструирование понятий. Вот ключевой фрагмент кантовской «Критики чистого разума» (Критики) по этому поводу:

«Философское познание есть познание разумом посредством понятий, а математическое знание есть познание посредством конструкции [из] понятий¹. Но конструировать понятие — значит показать а priori соответствующее ему созерцание. Следовательно, для конструирования понятия требуется не эмпирическое созерцание, которое, стало быть, как созерцание есть единичный объект, но тем не менее, будучи конструированием понятия (общего представления), должно выразить в представлении общезначимость для всех возможных созерцаний, подходящих

¹ Мы привели здесь дословный перевод фразы «...die mathematische aus der Konstruktion der Begriffe. Einen Begriff aber konstruieren, heißt: die ihm korrespondierende Anschauung a priori darstellen» (нем.), хотя есть и другой вариант перевода: «математическое знание есть знание посредством конструирования понятий» [КЧР, В741]. Эти переводы по-разному акцентируют специфику математической деятельности: «конструирование понятий vs конструкция [из] понятий», а в кантовской Критике присутствуют оба эти смысла. В первом случае под «конструированием» понимается процесс (акт конструирования), например проведение линии на чертеже, а во втором случае — результат этого процесса, созерцательная конструкция (нарисованный треугольник).

под одно и то же понятие... Единичная нарисованная фигура эмпирична, но тем не менее служит для выражения понятия без ущерба для его всеобщности, так как в этом эмпирическом созерцании я всегда имею в виду только действие по конструированию понятия, для которого многие определения, например величины сторон и углов, совершенно безразличны, и потому я отвлекаюсь от этих разных [определений], не изменяющих понятия треугольника» [КЧР, А713/В741¹]; выделено мной. — С. К.)

Это означает, что математика является не чисто интеллектуальной деятельностью, т.е. дискурсивной обработкой рассудочных понятий (символов), а задействует второй основной «ствол [человеческого] познания» [В30] — чувственность, который в случае математики выступает как схематизм чистых чувственных понятий, или трансцендентальная способность воображения. В термин «конструирование» Кант вкладывает следующий (необычный) смысл: конструировать понятие — значит сопоставить ему некоторое общезначимое созерцание, или кантовскую схему, реализацией которой в случае геометрии выступает чертеж. По сути дела, в своей модели математики Кант в общем виде описывает реальную геометрическую практику, которая предполагает работу не с самими математическими (геометрическими) понятиями (которые Кант именует «чистыми чувственными понятиями» [В180]), а с их «чертежными» (остенсивными) конструкциями — созерцаниями². Как замечает Кант, из понятия «треугольника» мы не можем извлечь какую-то существенную информацию о свойствах треугольника, например заключить, что сумма его углов составляет 180° , т.е. доказать соответствующую геометрическую теорему, но общезначимое созерцание (схема) треугольника, реализуемое посредством его чертежа, позволяет осуществить ряд математических

¹ Указание на страницы кантовской Критики [1] здесь и далее будем давать в стандартной международной пагинации [А (1-е изд.) // В (2-е изд.)].

² При этом надо учитывать отмеченный уже Платоном тот факт, что математик «работает» не с чертежом, а с представленной в нем «идеей» (resp. кантовским общезначимым созерцанием, или схемой): «Те, кто занимается геометрией, считают и тому подобным, предполагают в любом своем исследовании, будто им известно, что такое чет и нечет, фигуры, три вида углов и прочее... [И] когда они пользуются чертежами и делают отсюда выводы, их мысль обращена не на чертеж, а на те фигуры, подобием которых он служит [чертеж же является «образным подобием того, что можно видеть не иначе как мысленным взором» (там же)]. Выводы свои они делают только для четырехугольника самого по себе и его диагонали, а не для той диагонали, которую они начертили...» [Платон, Государство, кн. 6, 510с–511а; вставки наши. — С. К.]

действий по преобразованию треугольника в более сложную геометрическую конструкцию (за счет проведения дополнительных линий и т.п.) и доказать данную теорему, т.е. получить новое математическое знание. При этом схема треугольника [общезначимое созерцание] гарантирует, что полученный результат будет общезначимым, т.е. верен для всех (квантор всеобщности) треугольников, а не только для единичного (конкретного) треугольника чертежа.

Для прояснения своего тезиса о конструировании понятий Кант анализирует доказательство теоремы о равенстве суммы углов треугольника в [B744–5]). При этом он обращает внимание на то, что в ходе подобной математической деятельности исходное (рассудочное) понятие треугольника (по определению, как предмета, состоящего из трех углов) преобразуется в остенсивную созерцательную конструкцию — фигуру треугольника, что позволяет разложить его (ее) на составляющие: отрезки, прямые и углы, — осуществить дополнительное построение (проведение прямой через одну из его вершин и продолжение двух других сторон), а также использовать для доказательства выявленную при этом построении новую созерцательную информацию, например о равенстве соответствующих углов. За счет подобного введения новых объектов и, соответственно, новых действий с ними [в созерцательной среде], а также усмотрений открывающихся при этом новых очевидностей удастся «синтезировать» (получить) новое знание о треугольнике, т.е. доказать искомую теорему о том, что сумма углов произвольного (любого) треугольника равна 180 градусам.

Далее Кант распространяет данную трактовку на всю математику в целом, поскольку алгебру (алгебраические преобразования) можно также трактовать аналогичным — конструктивным — образом, хотя созерцательной «средой» здесь выступает уже не пространство [чертежа], а априорное время (алгебраические действия совершаются последовательно во времени). Вместе с тем можно показать, что кантовская модель математики с *mutatis mutandis* вполне применима и к современному абстрактному логико-математическому (по знанию (и даже к логике): см. наши работы [2, 3].

Кантовская модель математической деятельности как конструирования понятий, что предполагает спуск на созерцательный уровень чувственности и обратный подъем на «понятийный» уровень рассудка, представить достаточно просто¹.

¹ См. пояснения к схеме двухуровневой (рассудок и чувственность) математической деятельности в моих работах, посвященных кантовской [трансцендентальной] концепции [философии] математики [4, 5].

Тем самым человеческая математическая деятельность созерцательно-дискурсивна, поскольку задействует как рассудочное, так и образное мышление¹, в то время как развитие систем ИИ пошло по другому пути, заданному Д. Гильбертом и его последователями (формализм). Для Гильберта основой математики является аксиоматический метод, позволяющий представить математические теории как логические исчисления, «действия» в которых представляют собой уже чисто дискурсивные (рассудочные) процедуры, каковыми выступает логический вывод. Например, вместо геометрических (пространственных) построений мы могли бы, представив геометрию в качестве аксиоматической системы, заменить геометрические действия построением логического вывода. Такая формализованная математика уже чисто дискурсивна и может быть алгоритмизирована, т.е. представлена как логический вывод (по алгоритмам) в некоторой системе ИИ. Никакой созерцательной конструкции для совершения математических действий здесь уже не требуется: математические теоремы доказываются («выводятся») в соответствующей формальной системе. При этом следует различать собственно человеческую математическую деятельность как набор математических действий в созерцательных средах (пространства и времени) в нижней части схемы и ее логическое оформление в системах ИИ, которое может быть представлено в верхнем блоке схемы как формальный переход («вывод») от «формулы» A (например, системы аксиом) к «формуле» B ². И хотя гильбертовская программа формализации относительно всей математики (ее богатых разделов) потерпела крах: результаты ограничительных теорем К. Геделя, А. Тарского, А. Черча, А. Тьюринга, — тем не менее заданный ею тренд стал основополагающим для последующего развития систем ИИ, хотя при этом все выявленные ограничения для формальных систем здесь также действуют³.

¹ Специфику образного мышления мы обсудим ниже. В этой связи М. Мамардашвили как-то сказал, что «Сезанн мыслит [не понятиями, а] яблоками [т.е. образами]».

² Ср. с различием И. Лакатосом в его книге «Доказательства и опровержения» строгости [математического] доказательства и строгости [логического] анализа доказательства.

³ Обратим внимание на одну интересную альтернативу в современной математике/логике (теории вычислений), связанную с развитием так называемой «геометрической логики» («geometric logic»), в качестве модели геометрического мышления человека. Для ее формализации используется теория топосов, а семантически она может быть соотнесена с интуиционистской логикой (см., например, [6]).

В данной статье мы не ставим задачу обсуждения возможных способов преодоления ограничительных теорем (и развития систем ИИ); скромная цель нашего анализа здесь — показать принципиальное различие между математикой с «человеческим лицом» (модель которой была эксплицирована Кантом) и формализованной (алгоритмизированной) математической деятельностью, которые реализуют системы ИИ. Очевидно, что и естественный и искусственный разум «умеют» доказывать математические теоремы, а в последнее время они научились «работать» вместе, причем это приносит весьма значимые результаты (доказательство теоремы о четырех красках, теоремы Ферма), но при этом принципы их функционирования отличаются друг от друга: созерцательное «конструирование понятий» в математике остается прерогативой лишь человеческого разума¹.

* * *

2. Вместе с тем различие в осуществлении математической деятельности человека и ИИ, будучи показательным и важным, является лишь частным случаем более общего различия между естественным и искусственным интеллектом. Речь идет о феномене образного мышления (образного модуса сознания) человека, основой для которого выступают ментальные образы (*mental image*) нашего сознания. Точнее о проблеме эпистемологического статуса ментальных образов. С одной стороны, очевидно, что в нашем сознании существуют ментальные образы (греч. *phantasma*) как результат деятельности воображения или фантазии (*phantasia*; *φαντασία*): мы закрываем глаза и представляем образ стоящего передо мной стола, или вспоминаем его вчерашнее присутствие здесь, или воображаем подобный стол в другом месте, а наши сны представляют собой целые связанные последовательности таких «картинок». В человеческой культуре есть целая традиция создания и культивирования фантастических (вымышленных) образов типа русалок, кентавров или химер. Более того, художественные образы — основа художественного творчества, или искусства. Назовем такие ментальные картинки нашего сознания образами — 1. С другой стороны, можно ли считать наш процесс восприятия образным процессом, т.е. говорить о том, есть ли в нашем

¹ Хотя это не означает, что такой — человеческий — способ математического мышления является наиболее эффективным и подлежит обязательному моделированию в системах ИИ.

сознании образы вещей (которые назовем образами–2¹)? Вот что в этой связи пишет Э. Гуссерль в «Логических исследованиях» [7; приложение к §§ 11 и 20], где он подвергает уничижающей критике «[образную] теорию отражения» (Bildtheorie):

Против этого [теории отражения²] следует заметить, что концепция полностью упускает из виду самый важный пункт, а именно, что в акте представления как отражения на основе являющегося «объекта-образа» («Bildobjekt») мы имеем в виду отраженный (abgebildet) объект («основу образа» («Bildsujet»)). Однако образность объекта, функционирующего как образ, не есть, очевидно, [его] внутреннее свойство (не есть «реальный предикат»), как будто объект, как он, например, является шарообразным, является еще и образным. Каково же основание для того, что мы можем выйти за пределы данного лишь в сознании «образа» и отнести его в качестве образа к определенному объекту, который чужд сознанию? Указание на подобие между образом и вещью не продвинет нас дальше. Это подобие, по крайней мере если вещь действительно существует, без сомнения, имеет место как объективный факт. Однако для сознания, которое, как предполагают, имеет только образ, этот факт просто не существует; он не может служить для того, чтобы прояснить сущность представляющего, точнее, отражающего отношения к внешнему для этого отношения объекту (основе образа). Сходство между двумя предметами, пусть даже очень большое, не превращает один предмет в образ другого. Лишь способность представляющего Я воспользоваться подобным как образом-репрезентантом для подобного, иметь первое в качестве наглядно присутствующего и все же вместо него подразумевать другое, вообще делает образ образом... Такой неточный оборот речи, как «внутренний образ» (в противоположность внешнему предмету), нельзя допускать в дескриптивной психологии... Картина есть образ только для сознания, которое конституирует образ и которое впервые придает «значение» образа первичному

¹ Здесь образ дан в генетиве, т.е. как образ чего-то. Образ–2 является не самостоятельной сущностью, а лишь копией чего-то. Понятно, что образ–1 как результат работы воображения надо отличать от образа–2: соответственно, не любой образ–1 является образом–2.

² «... которая полагает, что в достаточной степени проясняет факт [наличия] акта представления... следующим образом: «вне» существует или существует по меньшей мере при определенных обстоятельствах сама вещь; в сознании существует, как ее представитель, образ» [там же; 7; с.392].

и являющемуся ему в восприятии объекту, благодаря своей апперцепции» [7; с.392–5]¹.

Тем самым образность является не собственным («реальным») предикатом имеющейся в сознании ментальной картинки (образ–1), а лишь мета–эпистемической характеристикой тех из представлений, для которых будет справедливо бинарное отношение образности «образ (как отражение чего-то) – оригинал». Образом–2, например, будет отражение предмета в зеркале, но лишь потому, что мы в этом случае выступаем такими сторонними наблюдателями, которые одновременно могут видеть и оригинал (предмет сам по себе) и его образ–2, и, соответственно, фиксировать бинарное отношение образности. В случае же наличия у нас внутренних ментальных образов (= образов–1) мы не можем выступать в роли стороннего наблюдателя, поскольку для этого у нас, помимо доступа к образу (посредством апперцепции или рефлексии), должен быть еще и прямой доступ (= восприятие) к оригиналу самой вещи, что как раз и не принимается теорией отражения. Поэтому, хотя имеющаяся у нас ментальная картинка и является образом–1 (как результатом деятельности воображения–образной силы), у нас нет достаточных оснований характеризовать ее в качестве образа чего-то [вещи], т.е. называть ее «отражением» внешней действительности, или образом–2.²

Далее гуссерлевскую критику развивает и дополняет Ж.–П. Сартр, который в своей работе «Воображение/Воображаемое» [8] развивает феноменологическую теорию воображения. Он отвергает традиционное учение об образах как копиях вещей и предполагаемое этим учение удвоение сущностей «вещь vs образ» (иллюзию имманентности), которое он связывает со склонностью обыденного рассудка населять наше сознание образами как «маленькими пододбиями вещей»³. Суть феноменологического подхода состоит в том, что есть не образы сознания [в сознании], а образный (фантазий-

¹ Здесь мы цитируем фрагмент сокращенно, без поясняющих скобок/вставок с немецкими терминами русского перевода.

² В дополнение к гуссерлевской критике Bildertheorie приведем еще один [известный] аргумент: если бы в нашем сознании были образы, то для их «видения» надо было бы ввести еще одни «глаза [ума]» и т.д. Гуссерль фиксирует это как бесконечную итерацию образов.

³ Ср.: «Введенные в заблуждение смещением предмета и психического содержания, упускают из виду, что предметы, которые нами «осознаются», не наличествуют в сознании просто как в коробке, так, что их можно было бы там обнаружить и взять, но что они прежде всего конституируются в различных формах предметной интенции как то, что они суть и что они значат» [7, с.156].

ный) модус сознания или особый образный интенциональный акт, который имеет свою специфику и принципиально отличается от акта чувственного восприятия (созерцания).

В середине XX в. серьезную экспериментальную и концептуальную (когнитивную) поддержку феноменологической критике традиционной теории воображения и тезису о без-образном характере нашего восприятия/познания оказал «экологический подход» к проблеме зрительного восприятия Дж. Гибсона [9]. Он основан на том факте, что человеческий глаз воспринимает не предмет (его образ), как это полагается классическими теориями восприятия, а модулированный световой поток (фотонов), имеющий свой особый источник (например, Солнце, но не вещь) и несущий «информацию»¹ в том числе и о воспринимаемом предмете, а имеющаяся у человека информационная система («тело – глаз – мозг») позволяет извлечь «информацию» в виде набора инвариантов. Соответственно, предметное (образное) восприятие выступает лишь феноменологическим описанием этого процесса, основанным на вольной аналогии (отождествлении) процесса восприятия и процесса передачи сообщения, которая в большей степени подходит для восприятия художественных образов (например, картин или рисунков), чем для восприятия предметов. В отличие от [информационного] сообщения в воспринимаемый нами световой поток информацию никто специально не вкладывает². Тем самым Гибсон предлагает/обосновывает не «геометрическую» (образную), а «алгебраическую» модель обработки (извлечения) информации как системы инвариантов, в рамках которой места для ментальных образов [восприятия] нет³. Причем если Гуссерль только сформулировал контраргумент о невозможности наличия образа – 2 в нашем уме, то Гибсон разрабатывает экспериментально обоснованную теорию без-образного восприятия⁴.

¹ Дж. Гибсон подчеркивает, что это не [количественная] информация в шенноновском смысле, модель которого применима, прежде всего, к передаче сообщений, а не к получению информации (в процессе восприятия) о внешнем мире.

² Как это предлагает, например Платон (Аристотель), наделяя вещи [идеальными] формами.

³ Здесь мы ограничились кратким изложением экологического подхода в качестве поддержки тезиса Гуссерля о без-образном характере нашего восприятия. Позиция же самого Дж. Гибсона более радикальна: он предлагает пересмотреть весь предшествующий концептуальный аппарат для описания процесса восприятия (в том числе и феноменологический).

⁴ В пользу тезиса Гуссерля – Гибсона говорит и то, что более простые существа (типа амёбы), извлекают информацию о мире и ориентируются в нем, но [психической] образной способностью, присущей «животным душам» (Аристотель) не обладают.

Тем самым налицо следующая апория (дилемма): наличие образов–1 vs отсутствие образов–2. Хотя эта апория не является антиномией: отрицание наличия в сознании образов–2 (образов вещей) не означает отрицания наличия в нашем сознании образов–1 (образов как таковых), а деятельность математики (геометрии) и искусства (живописи) является прекрасным подтверждением этого.

Заметим также, что новый импульс к обсуждению и поиску способов решения данной апории придала дискуссия по проблеме *mental image*¹ в середине 80-х гг. XX в, в ходе которой сформировалось два влиятельных противоборствующих лагеря, связанных с именами Ст. Косслина (признание образов) и З. Пилишина (их отрицание; алгебраическая модель мышления). Это обсуждение было напрямую связано с проблематикой «машинного разума» («следует ли системе AI наделять образным мышлением, имеющимся у человека, или нет?») и способствовала развитию когнитивных исследований и новых подходов по теме воображения².

Каким образом можно разрешить данную апорию образности? Суть нашего решения такова. [Тезис] Ментальные образы нашего сознания являются вторичными голограммами³, посредством которых воображение как способность по созданию образов визуализирует получаемую, по Гибсону, «информацию» о внешнем мире. [Расшифровка тезиса] Это означает, что создаваемые в ходе познавательного процесса ментальные образы не образуются на начальном этапе познания – восприятию, они не представляют собой, например, преобразованные воображением «картинки» вещей на сетчатке нашего глаза в собственно [ментальные] образы, которые затем «видит» наш «умственный взор» и «интерпретирует» соответствующим образом (распознавание образов) посредством понятий рассудка, как это постулируют восходящие к Аристотелю классиче-

¹ См., например: <http://plato.stanford.edu/entries/mental-imagery/>.

² К началу XXI в. в когнитивистике сложились три основных подхода к решению проблемы ментальных образов: 1) гипотеза двойного кодирования (Пэвио), 2) концептуально–пропозициональная гипотеза (Андерсен и Бауэр, Пилишин), 3) гипотеза функциональной эквивалентности (Шепард, Косслин). Подробнее см. [10, гл.10].

³ Поводом для данного тезиса послужила лекция Н. Хамфри «Изобретение сознания» (Москва, МГУ, 4.11.2015.; https://www.youtube.com/watch?v=ESvN_8pYA9Y), в которой был озвучен тезис о голографических эффектах нашего сознания (о концепции сознания самого Н. Хамфри см.: [11], https://en.wikipedia.org/wiki/Nicholas_Humphrey). Термин «голограмма» отсылает к математическому (космологическому) голографическому принципу (гипотеза Г. Хофта): вся информация, содержащаяся в некой области пространства, может быть представлена как «голограмма» – теория, которая «живёт» на границе этой области.

ские концепции репрезентации. Наши ментальные образы являются более поздним – в каком-то смысле эпифеноменальным – «изобретением» нашего сознания, или вторичной – геометрической – визуализацией извлекаемой первичной – алгебраической – информации как системы инвариантов (Гибсон). Голографический эффект подобной визуализации состоит в том, что 2-мерная алгебраическая (не-визуальная) информация преобразуется в геометрический (квази)3-мерный образ. В принципе, наше восприятие (= естественный интеллект) могло бы обойтись и без подобных голографических ментальных образов. Об этом говорит, например, опыт восприятия более «низких» живых существ, способности воображения у которых нет¹. Вместе с тем появившееся в ходе эволюции образное мышление человека, осуществляемое посредством воображения, обладает важной эпистемологической значимостью, поскольку позволяет переходить на мета-уровень и синоптически обзирать (осознание, ап-перцепция) получаемую информацию², что позволяет человеку сформировать целостный взгляд на мир (или осуществить рефлекссию)³.

* * *

3. В заключение нашего анализа коротко обратимся к третьей из заявленных нами тем – к теме сознания. Выше мы уже говорили о том, что важным различием искусственного и естественного интеллектов выступает феномен (само)сознания. Если естественный интеллект [человека] является сознательным, то системы ИИ сознанием не обладают и, более того, как показывает успешное современное функционирование ИИ, сознание не является необходимым компонентом интеллектуальной деятельности (мышления), т.е. сознание является, скорее, эпифеноменом нашего разума, дополняющим важнейшую для выживания в мире интеллектуальную деятельность. Соответственно, мы должны отказаться от тезиса о глубоком родстве интеллекта и сознания, рассматривая моделирование интеллекта и сознания как две разные задачи.

¹ По Аристотелю, воображение отсутствует на уровне «растительной души» (первый класс живого) и появляется лишь на уровне «животной души» (второй класс живого).

² Ниже мы соотнесем способность такого перехода на мета-уровень с сознанием [12]

³ Ср. с метафорой «левое (дискурсивное, логическое) vs правое (образное) полушарие».

Что собой представляет сознание человека?¹ Традиционно сознание (в философско-психологической традиции) соотносят с высшими функциями человеческого ума. Например, Кант мыслит свое «трансцендентальное единство апперцепции» как сущностное основание человеческого разума. Мы же склоняемся к тому, что человеческое сознание располагается гораздо ниже рассудка/разума. Например, Р. Коллингвуд соотносит сознание не с «высшим» рассудком, а со срединной способностью воображения, традиционно помещаемой между нашей чувственностью и рассудком². Сходную концепцию в настоящее время развивает Н. Хамфри, понимающий человеческое сознание как голографический эпифеномен нашего ума [11]. Это значит, что феномен сознания тесно связан с ментальными образами (см. п.2 выше), т.е. деятельностью воображения. Оба этих феномена представляют собой своеобразные голографические эффекты работы человеческого мозга. В этой связи можно говорить об «экране сознания», на котором располагаются ментальные образы и началом координат которого выступает наше «Я» (кантовская трансцендентальная апперцепция)³. Именно на этом апперцептивном экране сознания (ср. с декартовским «естественным светом разума») протекают мыслительные процессы: например, когда мы думаем над решением какой-то задачи (интроспекция и/или рефлексия); однако большинство собственно «вычислительных» (интеллектуальных, мыслительных) процессов протекают за пределами этого «экрана

¹ Подробнее о моем понимании феномена сознания (глобальной проблеме сознания) см. [12]. Там я определяю [универсальное] сознание как способность существа/системы выходить за пределы наличной ситуации, переходить на мета-уровень. Вместе с тем концепт сознания является исторически обусловленным и можно выделить по крайней мере пять исторических модусов сознания [см. мой доклад «Исторические модусы сознания» на VI конф. «Философия. Язык. Культура» (НИУ ВШЭ, 29.04.2015; (<https://www.youtube.com/watch?v=5iV54gOkJ8>), а также мое выступление на конференции памяти М. Фуко (МГУ им. М.В. Ломоносова, 29.10.2016)]. Например, концептуализация «сознания» в Античности/Средние века посредством понятия «души» (при том, что «душа» Платона – это не то же самое, что «душа» Аристотеля) отличается от «*cogito*» (классическая трактовка сознания) Декарта или «трансцендентального единства апперцепции» Канта, а психоаналитический концепт «подсознания» принципиально противоречит «естественному свету разума» Декарта (resp. эпохи Просвещения), т.е. классическому пониманию сознания как структуры «сознание + самосознание (апперцепция, рефлексия)».

² Выше мы уже говорили о тесной связи сознания и воображения.

³ Подробнее о кантовской модели сознания и ее возможном применении в области computer sciences см. мою работу [13].

сознания», на под-сознательном/бес-сознательном уровне (мышление vs сознание)¹.

* * *

Подведем итоги. Современные успехи «машинного разума» (ИИ) впечатляют. С некоторыми оговорками можно сказать, что искусственный интеллект (пусть и в слабом смысле) уже построен: проверку на тест Тьюринга построенный ИИ пройти сможет. Однако существующие системы ИИ не моделируют естественный интеллект человека. Соотношение между ними таково. Искусственный интеллект — это частный интеллект как набор отдельных вычислительных алгоритмов (программ) для решения определенного класса интеллектуальных задач. Естественный интеллект — это более сложная и цельная структура, включающая в себя не только собственно человеческий интеллект (кантовский рассудок), но и такие когнитивные компоненты (подструктуры), как образное мышление (на основе ментальных образов), феномен сознания/самосознания, рефлексии и др. Вместе с тем, возможно, нужно отказаться от создания сильного ИИ как точной копии человеческого разума и развивать ИИ следуя его собственной логике.

ЛИТЕРАТУРА

1. Кант И. Критика чистого разума. М.: Мысль, 1994.
2. Катречко С. Л. Трансцендентальный анализ математической деятельности: абстрактные (математические) объекты, конструкции и доказательства //Доказательство: очевидность, достоверность и убедительность в математике, М., КД «ЛИБРОКОМ», 2014а. с.с.86 – 120.
3. Катречко С.Л. Математика как «работа» с абстрактными объектами: онтолого-трансцендентальный статус математических абстракций //Математика и реальность. Труды Московского семинара по философии математики. М., Изд-во Московского университета, 2014б. с. 421–452 (см. также мои более ранние работы на эту тему: 1) Катречко С.Л. Моделирование рассуждений в математике: трансцендентальный подход //Модели рассуждений – 1: Логика и аргументация. Калининград: Изд-во РГУ им. И. Канта, 2007. с.63–90; 2) Катречко С.Л. Трансцендентальная философия математики //Вестник Московского университета. Серия 7 «Философия», № 2, 2008. М: Изд-во МГУ, 2008. с.88–106).
4. Катречко С.Л. Трансцендентальный анализ математики: абстрактная природа математического знания //Кантовский сборник, 2015, №2(52), с. 16–31 (<http://kant-online.ru/?p=3172>).

¹ В данном отношении представляется интересной описание собственной «бес-сознательной» интеллектуальной деятельности (математического творчества) А. Пуанкаре [14].

5. *Катречко С. Л.* Трансцендентальный анализ математики: конструктивный анализ математической деятельности //Кантовский сборник, 2016, №1(55), с.16–33 (<http://kant-online.ru/?p=3322>).
6. *Goldblatt, R.* Topoi, The Categorical Analysis of Logic. Amsterdam: Elsevier Science Publishers, 1984 (Ch.16 «Logical geometry»).
7. *Гуссерль Эд.* Логические исследования. Т. II(1). М.: Гнозис, ДИК, 2001.
8. *Сартр Ж.-П.* Воображаемое. Феноменологическая психология воображения. СПб, Наука, 2001.
9. *Гибсон Дж.* Экологический подход к зрительному восприятию. М., Прогресс, 1988.
10. *Солсо Р.* Когнитивная психология, СПб: Питер, 2006. (гл. 10).
11. *Humphrey N.* Soul Dust: The Magic of Consciousness, Princeton University Press, 2011.
12. *Катречко С. Л.* Проблема сознания как философская проблема // Философия сознания: Аналитическая традиция. Третьи Грязновские чтения. М.: «Современные тетради», 2009, с.128–134.
13. *Катречко С. Л.* Кантовская концепция сознания (познания) как модель «искусственного интеллекта» //Новое в искусственном интеллекте. Методологические и теоретические вопросы (под ред. Д. И. Дубровского и В. А. Лекторского). М.: ИИнтелЛ, 2005. С.137 – 140.
14. *Адамар Ж.* Исследование психологии процесса изобретения в области математики. (пер. с франц.). М.: «Советское радио», 1970.

ДИСКУССИЯ

DISCUSSION

讨论

Дискуссия состоялась на следующий день после проведения пленарного заседания. Мероприятием руководили представители кафедры философии и методологии науки философского факультета МГУ им. М. В. Ломоносова заведующий кафедрой, проф. В. Г. Кузнецов, в. н. с. А. Ю. Алексеев и доц. А. В. Чусов.

Дискуссию открыл доклад в. н. с. кафедры философии и методологии науки философского ф-та МГУ, доктора философских наук А. Ю. Алексеева на тему «КОМПЛЕКСНЫЙ ТЕСТ ТЬЮРИНГА». Поскольку философское изучение искусственного интеллекта было инспирировано концепцией А. М. Тьюринга (1950 г., «игра в имитацию»), позже названной «тестом Тьюринга» (в 1970-е годы), совершенно правомочно современную философию искусственного интеллекта начинать с тщательного анализа достижений в области тьюрингового тестирования, координации и комплексной интеграции тех исследований, в которых рассматриваются принципиальные вопросы: «может ли компьютер мыслить, понимать, творить, осознавать, любить, быть личностью и пр.?». Сегодня насчитывается свыше сотни частных тестов Тьюринга. При их обсуждении и происходит экспликация фундаментальных проблем искусственного интеллекта, когнитивных наук, нейронаук, электронной культуры. Комплексный тест Тьюринга выполняет следующие функции при анализе проблем электронной культуры: интеррогативную функцию – как правильно ставить вопросы об имитации, моделировании, репродуцировании

и рекреационировании когнитивных феноменов посредством компьютерных систем; дефинитную – об идентификации x-системы на предмет y-компетенций стороны z-наблюдателя с w-компетенциями; конструирующую – о разработке принципиального проекта компьютера, реализующего когнитивные феномены; критическую – всестороннее обсуждение «за» и «против» проекта искусственного интеллекта; конститутивную – задающую собственно человеческие, сверх-вычислимые параметры: свобода, ответственность, смысл, самость, Я и др. Исследования в области комплексного теста Тьюринга инспирируют формирование новой, постнеклассической парадигмы искусственного интеллекта, того, что можно назвать «концептуальным подходом» к ИИ. В этой парадигме совмещаются и комплексно применяются методы и средства логического, алгебраического, нейросетевого, семиотического, бионического и других подходов к ИИ. Собственно, на концептуальном уровне жесткое разделение методов на два подхода, скажем, на репрезентативный и арепрезентативный или на символичный и коннекционистский, представляются искусственной теоретической абстракцией. Такое разделение игнорирует практические инженерные подходы к определению вычислимости в смысле совмещения машины Тьюринга с машиной Корсакова.

Заслуженный профессор МГУ, доктор философских наук В.А. Кувакин проанализировал пространство внутреннего мира с целью отыскания новых ресурсов интеллектуальной и психологической безопасности человека в условиях беспрецедентного давления на него информации нормативного, перформативного, суггестивного и манипулятивного характера.

Х.Э. Мариносян, доктор философских наук, профессор, шеф-редактор журнала «Философские науки» отметил важную роль развития искусственного интеллекта для современной цивилизации. Вся предыдущая история человечества – это история окультуривания человека. Конечно, историю можно трактовать как развитие все более изощренных средств уничтожения людей, вплоть до средств самоуничтожения. Но всегда находились силы для того, чтобы одуматься и остановиться. Исследования искусственного интеллекта приводят к угрозе разрушения техногуманитарного баланса. Мы уже сейчас перешли некоторую черту, возможно, безвозвратно, когда естественный человек становится искусственным человеком. Существуют искусственные руки, сердце и пр. Неизвестно, каких успехов в создании полностью искусственного человека достигли исследователи в

каких-нибудь секретных лабораториях и центрах. Надо предостеречь людей. Назад к человеку! Как это сделать? Как разработать программу сохранения «души»? В этом-то и должны помочь философы. Первоначально, конечно, физики и математики начинали тему искусственного интеллекта. И отрадно, что сегодня эта тема обсуждается на философском факультете. Философское осмысление этого предмета, возможно, предотвратит самоуничтожение человека, и люди найдут в себе силы остаться людьми. Цивилизация будет спасена.

Данную тему продолжил В. И. Громыко, научный сотрудник МГУ имени М. В. Ломоносова. Он представил совместный доклад на тему «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ – НАСЛЕДНИК ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ». Соавторами доклада явились В. П. Казарян, доктор философских наук, профессор МГУ им. М. В. Ломоносова, Н. С. Васильев, доктор физико-математических наук, старший научный сотрудник МГТУ имени Н. Э. Баумана, А. Г. Симакин, старший преподаватель Российского университета дружбы народов и С. С. Аносов, главный специалист банка «Возрождение». Системно-информационная культура и постнеклассическая наука определились в ходе осуществления новой когнитивной революции, связанной с построением субъектом самого себя. Система образования должна проявить заботу о развитии способности учащегося к достижению этой цели, обусловленной предстоящей синтезирующей междисциплинарной деятельностью субъекта. Система искусственного интеллекта должна сыграть важную роль в деле самосоздания и самодетерминации человека. Эта задача образования может быть решена благодаря достигнутому уровню развития математической объективации на базе языка категорий и метаматематики. По мнению авторов, этот подход позволит решить классическую задачу искусственного интеллекта – превратить его в «пожизненного» партнера естественного интеллекта. Интеллектуальные инструментальные системы и суперкомпьютер смогут поддерживать пользователя, способствуя развитию смыслового, системного сознания на основе познавательных и творческих функций.

Э. А. Дейнека, координатор международных программ НСМИИ РАН, кандидат медицинских наук, г. Москва, в докладе «ЧТО НУЖНО РОБОТУ (И ЧЕЛОВЕКУ) ДЛЯ ПРОХОЖДЕНИЯ ТЕСТА АДЫ ЛАВЛЕЙС?» подчеркнула, что Алан Тьюринг (в статье 1950 г.) признает «аргумент леди Лавлейс» наиболее удачным «возражением» против тезиса о способности вычислительных машин имитировать разумность

человека. Аргумент звучит так: машины не могут творить, т.е. создавать то, что не детерминировано заложенным в них программным кодом-алгоритмом. Причем способны создать то, что человеку может показаться «шедевром», но не могут его «удивить» оригинальностью своего «творения». Вопрос «может/не может компьютер творить» можно было бы перефразировать так: может ли программист создать такой код, результат исполнения которого будет для него непредсказуем? Если ответить «да», значит, это не код (т.к. код – это алгоритм и, по определению, результат его исполнения полностью предсказуем). Если ответить «нет», значит, это не программист (т.к. программист, по определению, должен уметь формализовать любую задачу и алгоритмизировать ее решение). Однако под творчеством понимается научное творчество или художественное? Ведь между ними есть принципиальное различие. Научное творчество направлено на отыскание оптимальных способов пребывания человека в адекватном отношении с физической и объективной социальной реальностью (решение практических задач). Художественное творчество направлено на отыскание оптимальных способов пребывания человека в адекватном отношении с теми состояниями, к которым его приводит (периодически возникающая, временная) невозможность пребывания в адекватном отношении с физической и объективной социальной реальностью (фантазирование). Далее Э.А. Дейнека представила программу «искусственного поэтического интеллекта», который «творит» весьма интересные и глубокие «сонеты» путем рекомбинации строчек-стихов в каждом сонете по схеме классической сонетной рифмовки (aBaB aBaB ccD eeD) и привела пример весьма высокохудожественных сонетов. В этих сонетах использовалось смешение литературной и жаргонной, сленговой лексики; присутствовало обилие многозначных слов, абстрактных понятий, мифических аллегорий, метафор, образных выражений; применялись омонимы (технически для рифм), идиоматические обороты, «избитые фразы», плеоназмы («спуститься вниз», «увидеть собственными глазами» и т.п. – семантическая избыточность, создающая иногда комический эффект: «она отнюдь совсем не глупа») для сглаживания возможных смысловых нестыковок при комбинировании строк-стихов, благодаря имитации соответствующей стилистики в целом (в том числе внутри строк, а не только между строками-стихами), оксюмороны (сочетание несочетаемого: «жгучий холод», «холодный жар», дающее иногда новые концепты) – тоже для сглаживания возможных

нестыкков. Как объяснить такие поэтические способности машины? В итоге получается стилистика «пророческого» типа, напоминающая катрены Нострадамуса или предсказания Дельфийского оракула. В итоге человек, слушающий такие сонеты, должен иметь чутье на схватывание целостного смысла подобного произведения, если, конечно, перед ним стоит задача отличить произведения машины-поэта от произведения человека-поэта. Часто такая задача не возникает.

Практические применения теста Тьюринга представили в совместном докладе «СЧАСТЬЕ – ЭТО КОГДА ТЕБЯ ПОНИМАЮТ: ПОЛЕЗНЫЙ НЕЧЕЛОВЕЧЕСКИЙ ИНТЕЛЛЕКТ» А. Р. Ефимов, кандидат технических наук, директор робототехнического центра Сколково, г. Москва и О. С. Аврясова, руководитель направления этого центра. Докладчики отметили, что тест Тьюринга – это не только теоретическая задача исследований искусственного интеллекта. Сегодня проект тьюринговой игры в имитацию звучит как идея проверки «человечности» собеседника, которая предполагает общение человека с другим человеком и с компьютерной программой в течение пяти минут, только в текстовом режиме. Если компьютер сможет обмануть как минимум 30 % собеседников, тест считается пройденным. Эксперты общаются одновременно с живым человеком и роботом, находясь в разных комнатах и не видя друг друга. После завершения теста каждый из них должен сказать, кто из двух его собеседников был человеком, а кто – программой. Несомненным достижением российских программистов в области искусственного интеллекта, как полагают докладчики, является выигрыш чат-бота «Евгений Густман» 7 июня 2014 года на соревнованиях, организованных университетом Рединга в честь 60-ой годовщины со дня смерти А. М. Тьюринга. 33 процента судей после пяти минут общения признали этого «мальчика из Одессы» за человека. В соревновании принимали участие 5 чат-ботов и 30 судей. Каждый судья общался с каждым из пяти чат-ботов и с человеком. Однако русские ли впервые прошли тест Тьюринга? А. Ю. Алексеев категорически против такого заявления (см. статью в настоящем Сборнике).

В. А. Аршинов, доктор философских наук, профессор, главный научный сотрудник Института философии РАН представил работу «ПРОБЛЕМА СЛОЖНОСТИ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ». Сложность в ее современном, постнеклассическом понимании в эпистемологическом отношении не редуцируется к объективному или субъективному своему измерению. Сложность – это

разворачивающаяся во времени становящаяся сеть эмерджентных событий, контингентно определенных по отношению к прошлому и столь же неопределенных по отношению к будущему. Возможно то, что постнеклассическая сложность обладает «квантовоподобными свойствами». Концепты «наблюдатель», «наблюдаемость», «контингентность», «контекстуальность» образуют рекурсивно сцепленный понятийный кластер. Поэтому следует ввести «наблюдателя темпоральной сложности». Его осмысление ориентирует на поиски путей преодоления декартовского разграничения протяженной и непротяженной субстанций, на включение сознания в контекст понимания проблематики постнеклассической сложности как ключевого междисциплинарного понятия. Такой подход не исключает понимания сложности, основанного на понятии алгоритмической сжимаемости. Он предполагает наблюдателя, распознающего регулярности, паттерны в представленных ему последовательностях чисел или иных символов. Сложность оказывается релятивной по отношению к наблюдателю, точнее, к множеству наблюдателей, их точек зрения, перспектив, контекстов и интеракций. В этой связи особый интерес представляет применение постнеклассической концепции наблюдателя к наблюдателю теста Тьюринга.

В. Л. Васюков, доктор философских наук, заведующий кафедрой истории и философии науки Института философии РАН, г. Москва, в докладе «ВИТГЕНШТЕЙНИАНСКИЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ» продолжил тему, которая неоднократно звучала на семинарах НСМИИ РАН, по поводу проекта «Интенциональный компьютер». На этот раз он предложил использовать идеи Л. Витгенштейна и тем самым преодолеть тупики искусственного интеллекта, основанного на идеях М. Хайдеггера. Модифицируя понятие интенциональности, данное Дж. Серлом, докладчик представил гипотетическую модель витгенштейнианского интенционального искусственного интеллекта, основанную на комбинациях логических и дедуктивных систем.

Д. В. Зайцев, доктор философских наук, профессор кафедры логики философского факультета МГУ имени М. В. Ломоносова, и Н. В. Зайцева, кандидат философских наук, профессор кафедры гуманитарных и социальных наук, профессор Всероссийской академии внешней торговли, г. Москва, в совместном докладе «ИНТЕНЦИОНАЛЬНОСТЬ, АППРЕЗЕНТАЦИЯ И КАТЕГОРИЗАЦИЯ» также частично затронули тему интенциональности. На наш взгляд, докладчики исходили из классической логической традиции.

Интенциональность – это понятие, фиксирующее направленность человеческого сознания на какой-нибудь предмет. Аппрезентация – это перенос смысла с предмета (образца) на новый случай на основании сходства этого предмета. Категоризация – это подведение феномена под определенную рубрику опыта, категорию и признание его членом этой категории. Категоризация проходит два этапа: типизация чувственных данных по отдельности как признаков, сторон, характеристик предмета и, во-вторых, узнавание/распознавание этого единства как предмета, на который направлено познание. Собственно, на последнем этапе и появляется ключевая роль интенциональности в формировании интеллекта как системы понятий, категоризирующих явления. Современная теория понятия (Е. К. Войшвилло, В. А. Бочаров, В. И. Маркин), представляющая собой формальное представление понятий, отношений между ними и операций над понятиями, не адекватна концепции интенциональности. Поэтому предлагается вариант функциональной теории понятия, однако, в отличие от классической фрегевской концепции (понятие как предметно-истинностная функция), докладчики предложили оригинальный подход: понятие как релевантная функция, представимая в формате лямбда-исчисления. То есть интенциональность – это функция между стимулами и интенциональными (осмысленными) объектами, релятивизированными относительно субъекта. Далее докладчики привели примеры лямбда-формализации таких особенностей интенциональности, как типизация сторон предмета и категоризация.

Еще одну работу на логико-эпистемологическую тему предложил А. М. Анисов, доктор философских наук, ведущий научный сотрудник сектора теории познания Института философии РАН, г. Москва. В работе «КОМПЬЮТЕРНЫЕ МОДЕЛИ ВРЕМЕНИ» подчеркивается, что некоторые свойства времени позволяют представлять их геометрическими, т.е. пространственными средствами. Однако ошибочно сводить феномен времени к пространству, как это фактически делается в современной физике с её теорией единого пространства-времени. Время, полагает докладчик, суть вычислительный процесс. Переход от настоящего к прошлому и будущему – это результат вычислений, производимых самой природой. Поэтому в качестве базовых средств представления времени должны использоваться именно компьютерные модели. Однако современные теории вычислимости здесь не очень-то пригодны. Поэтому нужны исследования времени как феномена, обладающего вычислимым статусом.

Ю. Ю. Петрунин, доктор философских наук, профессор кафедры стратегических коммуникаций факультета государственного управления МГУ имени М. В. Ломоносова, г. Москва, в докладе «LAPIS PHILOSOPHORUM, ARTIFICIAL INTELLIGENCE И PERPETUUM MOBILE: ЭВОЛЮЦИЯ ВЗАИМООТНОШЕНИЙ МИФА И НАУКИ» акцентировал внимание на противоречивости того явления, которое принято называть искусственным интеллектом, и рассмотрел роль философии в исследованиях искусственного интеллекта.

И. Ф. Михайлов, кандидат философских наук, научный сотрудник Института философии РАН, г. Москва, представил тему «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ НАУКИ: ПЕРСПЕКТИВА АНТИРЕПРЕЗЕНТАЦИОНАЛИЗМА». Он проследил параллели в историческом и проблемном развитии теории искусственного интеллекта и когнитивных наук. Вместе с тем, как отметил автор, наблюдается попытка обосновать антирепрезентационализм как наиболее перспективную позицию в обоих научных направлениях.

С. Л. Катречко, кандидат философских наук, доцент школы философии факультета гуманитарных наук Национального исследовательского университета «Высшая школа экономики», г. Москва, в докладе «ЕСТЕСТВЕННЫЙ VS ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ФЕНОМЕН СОЗНАНИЯ» выделил три принципиальных отличия человеческого интеллекта от существующих систем ИИ: 1) математическое мышление человека; 2) наличие ментальных образов в человеческом уме и отсутствие (необязательность) моделей как таковых в системах ИИ; 3) феномен сознания, без которого нельзя помыслить человеческий ум, но который отсутствует (или является необязательным компонентом) в проектах существующих систем ИИ.

А. А. Кобляков, профессор, декан композиторского факультета Московской государственной консерватории имени П. И. Чайковского, руководитель междисциплинарного семинара «Музыка и наука», г. Москва, в докладе на тему «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ТВОРЧЕСТВО» отметил, что функции творчества расширяют наши представления об ИИ. Для моделирования творчества необходима объяснительная модель. В этом качестве предлагается модель музыкального произведения, обобщенная на основе методологии нового типа – т.н. «проблемно-смыслового подхода». Единицей смысла является вопрос/ответная структура (в музыке – диада «неустой/устой», базовая ячейка ладового тяготения). И само произведение, и наше восприятие имеют изоморфную иерархическую

структуру, в основе которой – ожидание ответа на вопрос. Как результат – формируется нелокальная сверхсегментная интегрирующая связь, что считается базовой проблемой в теории ИИ. Модель произведения допускает обобщение в логической форме («логика творчества» в данной терминологии), особенностью которой являются отношения нового типа («трансмерные отношения» в данной терминологии). Выявляется креативная роль отрицания в творческом процессе, что необходимо учитывать уже в самом начале программирования ИИ нового поколения. Эстетическое моделирование помогает лучше понять и деятельность мозга. Основу его деятельности составляет все та же вопрос/ответная структура, первичная ячейка смысла (=мысль). Возникает возможность описания разных функций и работы мозга в целом однокорневым образом (через постоянный диалог «вопрос-ответ»). Итак, музыкальный шедевр предложен как эталонная модель, объясняющая сопряжение значения и смысла, как универсальная модель становления осмысленного целого на любом субстрате.

Н. М. Смирнова, доктор философских наук, профессор, главный научный сотрудник Института философии РАН, заведующая сектором философских проблем творчества, руководитель Всероссийского семинара «Философия творчества», г. Москва, в докладе «ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ФЕНОМЕНОЛОГИЯ: ВЗАИМОДОПОЛНИТЕЛЬНОСТЬ ИССЛЕДОВАТЕЛЬСКИХ СТРАТЕГИЙ» обозначила проблемы философии искусственного интеллекта в предметных срезах философии и когнитивных наук и выявила эвристический потенциал феноменологии в исследовании ряда проблем философии искусственного интеллекта. Она показала, что, несмотря на различие исходных когнитивных презумпций, феноменология способна внести существенный вклад в решение важных проблем философии искусственного интеллекта, к примеру, онтологии чистой возможности и ментальной каузальности.

Т. Б. Кудряшова, руководитель Ивановского отделения НСМИИ РАН, доктор философских наук, заведующая кафедрой философии Ивановского государственного химико-технологического университета, в докладе «О РОЛИ ПАМЯТИ В УСЛОВИЯХ ЭЛЕКТРОННОЙ КУЛЬТУРЫ!» подчеркнула, что в современном, насыщенном информационными технологиями мире, часто можно услышать предположения о том, что дальнейшее развитие общества будет связано с разделением функций между интеллектуальными системами и

человеком. Предполагается, что человек будет исполнять функции, требующие нетривиального, творческого решения. А интеллектуальные системы будут исполнять более простые функции, подвергающиеся алгоритмизации, а также рутинные операции, к числу которых ошибочно относят запоминание, хранение информации и ее извлечение в нужный момент. Такая установка опасна тем, что, отдав функцию запоминания машине, человек рискует вместе с этим потерять свою способность к творчеству, поскольку усилия по вспоминанию являются неотъемлемой составляющей творческого процесса. Сопоставление феноменологии памяти и феноменологии воображения позволяет сделать вывод, что в творческом процессе оба феномена тесно взаимодействуют, а роль памяти вовсе не сводится к пассивному припоминанию прошлого. Подлинным условием творчества является интеллектуальное усилие по вспоминанию, активизирующее различные планы сознания. Таким образом, сохранение ведущей роли человека в сложных интеллектуальных технологических системах подразумевает развитие не только воображения, мышления и прочих когнитивных способностей, но и способности к активному вспоминанию как части творческого и продуктивного действия. Это особенно важно подчеркивать в образовательном процессе, мотивируя студентов к запоминанию текстов, идей, а не их механическому воспроизведению через чтение с листа.

И. В. Мелик-Гайказян, доктор философских наук, профессор, заведующая кафедрой истории и философии науки Института теории образования Томского государственного педагогического университета, г. Томск, в докладе «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СЕМИОТИЧЕСКАЯ ДИАГНОСТИКА АТТРАКТОРОВ КОНВЕРГЕНТНЫХ ТЕХНОЛОГИЙ» отметила связь исследований искусственного интеллекта с исследованиями НБИКС. Следует отметить, что отношения между синергетикой и семиотикой установлены докладчицей на основе ее оригинальной трактовки феномена информации.

Е. А. Никитина, координатор молодежных проектов НСМИИ РАН, доктор философских наук, профессор кафедры философии, социологии и политологии Московского технологического университета (МИРЭА), г. Москва, в докладе «ФИЛОСОФСКО-МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ РАЗВИТИЯ ИНТЕЛЛЕКТУАЛЬНОЙ РОБОТОТЕХНИКИ» провела анализ философско-методологических проблем естественного интеллекта, социально-философских аспектов развития интеллектуальной робототехники, философско-психологических проблем

взаимодействия человека и робота, актуальных проблем т.н. «роботики». Также ею была обоснована необходимость социально-гуманитарной экспертизы научно-технических проектов с участием робототехнических систем.

О.Э. Петруня, кандидат философских наук, доцент кафедры философии и социальных коммуникаций Московского авиационного института (национального исследовательского университета), г. Москва, в докладе «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ЧЕРЕЗ ПРИЗМУ ДИМЕНСИОНАЛЬНОЙ ОНТОЛОГИИ» представил т.н. антропный принцип (критерий человекомерности) исследований ИИ. Для реализации этого принципа методологически плодотворной представляется дименсиональная онтология В. Франкла. Постнеклассический критерий человекомерности, по мнению докладчика, оказывается продуктивным не только при решении задач создания систем адаптации человека к технике, но и при создании самой вычислительной техники.

Р.Г. Ершова, доктор психологических наук, профессор Государственного социально-гуманитарного университета, руководитель постоянно действующей международной конференции по проблемам электронной культуры «Цифровое общество как культурно-исторический контекст развития человека», г. Коломна, в докладе «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ПСИХОЛОГИЧЕСКИЙ ВЫЗОВ» рассмотрела две группы вызовов, связанных с развитием современных цифровых технологий, в том числе связанных с искусственным интеллектом. Имеются актуальные последствия этого «технологического прорыва»: социальные, социально-психологические, психологические, нейропсихологические. Наблюдаемые и прогнозируемые изменения неоднозначны, поэтому необходимы глубокие исследования проблем электронной культуры.

А. И. Панов, кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН, г. Москва, в докладе «БИОЛОГИЧЕСКИ И ПСИХОЛОГИЧЕСКИ ПРАВДОПОДОБНЫЕ МЕТОДЫ МОДЕЛИРОВАНИЯ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ» полагает, что целью ИИ является построение правдоподобных гипотез (моделей) работы когнитивных функций человека, которые бы согласовывались как с экспериментальными данными когнитивных психологов, так и с нейрофизиологическими данными о строении и функционировании мозга человека. Поэтому перспективно использование

культурно-исторического подхода и психологической теории деятельности (Лурия, Выготский, Леонтьев) как основы описания работы когнитивных функций, применение идей прикладной семиотики (Поспелов, Осипов) и использование моделей работы кортикальной колонки (Маунткэсл, Хокинс) и нейронного ансамбля (Эйдельман, Иваницкий) в качестве нейрофизиологического «заземления» проекта. Далее докладчик раскрыл особенности этого интегрального проекта и показал, как в нем осуществляется моделирование внимания, продуцирование нового знания (концепта), планирование поведения, построение картины мира субъекта на основе текстов, генерация сообщений на основе картин мира определенного типа (т.н. «виртуальные ассистенты»).

В.Г.Редько, доктор физико-математических наук, профессор, заместитель директора Института оптико-нейронных технологий РАН, г.Москва, свой доклад «МОДЕЛИРОВАНИЕ КОГНИТИВНОЙ ЭВОЛЮЦИИ» начал с фундаментального вопроса гносеологии: почему логический вывод, сделанный человеком, можно использовать в познании природы? Докладчик полагает, что исследовать происхождение логического мышления, интеллекта человека можно путем построения математических и компьютерных моделей когнитивной эволюции, которые прослеживают весь путь биологической эволюции от простейших до человека и позволяют четко показать, как на этом пути возникали способности познания закономерностей природы. Такие компьютерные модели разрабатываются в рамках проекта «Искусственная жизнь». Этот проект направлен на экспликацию формальных принципов организации биологической жизни и часто исследует более общие формы жизни, чем те, которые существуют на Земле, т.е. изучается жизнь, какой она могла бы в принципе быть («life-as-it-could-be»), а не обязательно та жизнь, какой мы ее знаем («life-as-we-know-it»). Докладчик показал принцип создания т.н. «когнитивной архитектуры», возможность ее программирования посредством мультиагентных систем и продемонстрировал моделирование «когнитивных способностей ворон».

А.В.Савельев, заместитель главного редактора журнала «Нейрокомпьютеры: разработка и применение», г.Москва, в докладе «ИНЖЕНЕРНЫЙ ПОДХОД В МЕТОДОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» привел обзор исторических методологических аспектов нейрокомпьютинга как постнеклассической парадигмы искусственного интеллекта, обсудил терминологические проблемы

определения нейрокомпьютинга в сильном и традиционно слабом его понимании, вопросы их адекватности, роль коннекционизма. Докладчик проанализировал бионические вопросы моделирования с помощью инженерных методов, связь парадигмы нейрокомпьютинга с моделированием социальных процессов, а также возможные расширения нейрокомпьютинга и некоторые его футурологические перспективы. На основе убедительного демонстративного материала докладчик проследил возможный путь от аналоговых компьютеров к нейрокомпьютеру будущего. Первый аналоговый компьютер был создан в России под руководством академика Алексея Николаевича Крылова в 1904 г. для моделирования качки корабля. Далее СССР занимал лидирующее положение в индустрии выпуска аналоговых компьютеров до 1970-х годов.

И. В. Степанян, доктор биологических наук, кандидат технических наук, ведущий научный сотрудник Института машиноведения РАН им. А. А. Благонравова, ведущий научный сотрудник МГК им. П. И. Чайковского, научный руководитель ЗАО «Интеллект» г. Москва, представил доклад на тему «ПРОЕКТ 'СЕМАНТИЧЕСКИЙ МОЗГ'». Докладчик проанализировал значительные объемы трудов ряда выдающихся писателей и деятелей культуры методом нейросемантической кластеризации. Для каждого автора построены индивидуальные карты смыслов различной глубины детализации, в каждой из которых проект «Семантический мозг» направлен на приложения нейросетевой технологии к философии творчества и философии мысли в целом. Ранее докладчик получил результаты по применению технологии нейросемантической кластеризации при анализе текстов и другой информации, включая биоритмы человека. Ключевой особенностью проекта «Семантический мозг» является перевод одномерных линейных текстов в двумерное пространство – карту смыслов. Это особый объект с богатыми математическими свойствами. Одна из ключевых идей проекта «Семантический мозг» – возможность изучения мозга человека, включая особенности психики, с использованием нейросетевой технологии семантической кластеризации значительных объемов информации, отраженной и сохранившейся в письменных трудах. Разработанный инструментальный нейросетевой кластеризации был применен для анализа трудов различных авторов. Проиндексированы сочинения ряда выдающихся писателей и деятелей науки, культуры, политики. Для каждого автора построены семантические карты как по отдельным трудам, так

и по всему творчеству в целом. В результате проведенного нейросемантического анализа были выявлены общие и различные для отдельных авторов характерные особенности. В частности, было установлено, что многие фразы и выражения в рукописях Н. В. Гоголя в значительной степени отличаются от фраз и выражений в рукописях других классических писателей. В то же время, у части авторов в центральном кластере расположено слово «Я», чего не наблюдается у другой части авторов. Эти и другие особенности отображены в полученных графических смысловых картах, доступных для быстрого визуального анализа.

А. В. Чусов, кандидат философских наук, доцент кафедры философии и методологии науки философского факультета МГУ имени М. В. Ломоносова, г. Москва, в докладе «СФЕРА ИССЛЕДОВАНИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА КАК ОБЪЕКТИВАЦИЯ» предложил единый подход к рассмотрению гетерогенных систем, включающих человека, сделав акцент на онтологических аспектах превращения исследований ИИ в сферу деятельности. Исходя из задачи объяснения того, как в мире появляются субъекты, он указал на параллели между возникновением таких сфер человеческой деятельности, как экономика и исследования ИИ. Они существенно различаются тем, что в последней использование частных результатов отдельных исследователей (локальные обмены) в общем воспроизводится, но спорадически и несущественно.

Е. А. Янковская, координатор образовательных проектов НСМИИ РАН, кандидат философских наук, ведущий научный сотрудник Российского экономического университета имени Г. В. Плеханова, г. Москва, в докладе «ГЕТЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ ЗНАНИЙ» отметила, что традиционные трактовки коннекционизма не учитывают сложность базовых элементов, объединяемых в сеть. Гетерархическая организация знаний позволяет моделировать континуально сложные элементы системы знаний.

М. А. Шестакова, кандидат философских наук, доцент кафедры философии и методологии науки Московского государственного университета им. М. В. Ломоносова, г. Москва, в докладе «МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ МНОГОАГЕНТНЫХ СИСТЕМ» показала, что изучение многоагентных систем представляет собой широкую междисциплинарную область исследований. Одной из актуальных задач является включение отдельных аспектов конфликтологии в теорию и методологию этих систем. Она усматривает

связующее звено между конфликтологией и теорией многоагентных систем в принципе ограниченной рациональности и ставит вопрос о перспективах исследования ресурсных конфликтов в рамках этих систем.

Руководители дискуссии:

***А. Ю. Алексеев
В. Г. Кузнецов
А. В. Чусов***

ЗАКЛЮЧЕНИЕ CONCLUSION

结论

В большинстве докладов затрагивались методологические проблемы взаимосвязи различных дисциплинарных областей, которые возникают при исследовании и проектировании систем ИИ: философии, биологии, психологии, лингвистики, социологии, вычислительной техники, информатики, программирования, робототехники, математики и др. Поэтому не случайно базовым организатором конференции явилась кафедра философии и методологии науки философского факультета МГУ имени М. В. Ломоносова.

В целом основную проблематику, которая была озвучена в пленарных докладах на конференции, можно разделить на две большие группы. Первая группа проблем была сфокусирована на методологических вопросах, касающихся взаимодействия различных научных дисциплин в исследованиях искусственного интеллекта. Прежде всего, рассматривался вопрос о роли философии в таких исследованиях. Вторая группа проблем касалась философского осмысления различных аспектов разработки искусственных интеллектуальных систем.

Выделим следующий перечень проблем.

1. Проблема прохождения комплексного теста Тьюринга.
2. Проблема философского анализа конвергентных технологий.
3. Проблема моделирования реальных психологических процессов мышления, в том числе тех, которые не поддаются однозначной формализации.
4. Проблема использования суперкомпьютеров для конструирования систем искусственного интеллекта.
5. Проблема влияния искусственных интеллектуальных систем на культурные феномены.

6. Развитие проблематики искусственного интеллекта как фактор развития электронной культуры.

7. Проблема влияния искусственного интеллекта на антропологическую идентичность и человеческое мышление.

8. Проблема уточнения концепций естественного интеллекта.

9. Проблема изучения механизмов, лежащих в основе разумного поведения.

10. Проблема критерия демаркации искусственное/естественное.

11. Проблема взаимосвязи когнитивной науки и феноменологии при исследовании естественного интеллекта и его моделирования в искусственных интеллектуальных системах.

12. Проблема этических аспектов разработки искусственного интеллекта, который способен превзойти своего создателя.

13. Проблема компьютерного моделирования социальных процессов с использованием многоагентных систем.

14. Проблема изучения искусственного интеллекта как сложной самоорганизующейся системы.

15. Необходимость новой когнитивной революции в когнитивной науке.

16. Необходимость принципиально новых идей для создания искусственных когнитивных систем.

17. Проблема биологоподобного искусственного интеллекта, для которого необходима физикалистская теория естественного интеллекта, решающая фундаментальный вопрос о том, откуда из операций нервной системы берутся свойства разума и сознания.

18. Проблемы интеллектуализации робототехнических систем.

Главную роль в решении этих проблем, в особенности в их идентификации и постановке, играют специалисты в области философской методологии междисциплинарных исследований. Они призваны обеспечить систематизацию, унификацию, кодификацию, формализацию, координацию и, наконец, интеграцию методов и средств решения таких проблем и в частных прикладных исследованиях проектов интеллектуальных систем, и в современных глобальных и общенациональных проектах, таких как «геном человека», «искусственный коннектом», «умные облака».

Специальная издательская серия «Философия искусственного интеллекта»

Философия искусственного интеллекта – широкая область современных междисциплинарных исследований, в которой ставятся в центр внимания общетеоретические, методологические и мировоззренческие вопросы изучения, имитации, моделирования и репродуцирования когнитивных феноменов. От решения этих вопросов в существенной мере зависит формирование новых прорывных направлений в изучении деятельности головного мозга, сознательных и бессознательных явлений человеческой психики, развитие информационных технологий и всего комплекса НБИКС-конвергенции: нанотехнологий, биотехнологий, информационных, когнитивных, социальных технологий и соответствующих им областей научного знания.

Дискуссии по философским проблемам искусственного интеллекта постоянно отражаются в известных отечественных периодических изданиях: «Вопросы философии», «Философские науки», «Философия и методология науки и техники», «Журнал высшей нервной деятельности имени И. П. Павлова», «Нейрокомпьютеры: разработка и применение» и др. Тем не менее этих издательских средств недостаточно для охвата постоянно расширяющейся сферы философско-методологических исследований искусственного интеллекта. Поэтому НСММИ РАН в 2005 году учредил собственное издательство «ИИнтелл».

За годы работы Совета было издано несколько десятков монографий и сборников трудов конференций.

Так, первая книга вышла в издательстве Института философии РАН: «Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г.» – М.: ИФ РАН, 2005. – 400 с.

В последующем основная литература нашего Совета стала издаваться в ряде издательств. Так, в издательстве «ИИнтелл» в 2005 году была издана первая книга Совета: «Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д. И. Дубровского и В. А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с.». В следующем, 2006 году вышла первая крупная монография по современным проблемам искусственного интеллекта: «Искусственный интеллект: междисциплинарный подход. Под ред. Д. И. Дубровского и В. А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с.». В этом же году вышел сборник работ первой молодежной конференции: «Искусственный

интеллект: философия, методология, инновации. Материалы Первой Всероссийской конференции студентов, аспирантов и молодых учёных, г. Москва, МИРЭА, 6–8 апреля 2006 г. Под ред. Д. И. Дубровского и Е. А. Никитиной – М.: ИИнтелЛ, 2006. – 440 с.».

В последние годы издательство «ИИнтелЛ» специализируется на выпуске книг, отражающих работу секций Совета. Материалы секции «Электронная культура» нашли отражение в книгах «Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А. Ю. Алексеева, С. Ю. Карпук – М., 2009» и «Электронная культура: феномен неопросветительства. Под ред. А. Ю. Алексеева, С. Ю. Карпук – М.: ИИнтелЛ, 2010. – 228 с. – 260 с.».

В 2014 году вышла книга «150 лет 'Рефлексам головного мозга'. Сборник научных трудов, посвященных изданию статьи И. М. Сеченова (23 ноября 1863 г.)» / Отв. ред. А. Ю. Алексеев, Ю. Ю. Петрунин, А. В. Савельев, Е. А. Янковская. – М.: ИИнтелЛ, 2014. – 432 с. Этот сборник стал первой книгой в серии «Нейрофилософия» (гл. редактор серии – проф. Д. И. Дубровский). С 2015 года под этой рубрикой стали выпускаться ежегодники «Актуальные вопросы нейрофилософии». В них отражается работа Международного междисциплинарного семинара «Нейрофилософия», который ежемесячно проходит на базе философского факультета МГУ им. М. В. Ломоносова.

Профессор Н. М. Смирнова, заведующая сектором философских проблем творчества, является главным редактором серии «Философия творчества». На сегодняшний день вышло два тома одноименного с серией Ежегодника: «Философия творчества: материалы Всероссийской научной конференции, 8-9 апреля 2015 г., Институт философии РАН, г. Москва / Под ред. Н. М. Смирновой, А. Ю. Алексеева. – М.: ИИнтелЛ, 2015. – 452 с.» и «Философия творчества. Ежегодник / РАН. ИФ. Сектор философских проблем творчества; Ред. кол.: Смирнова Н. М., гл. ред. и др. – М., 2016. – Выпуск 2, 2016: Когнитивные и социокультурные измерения / Ред. Смирнова Н. М., Майданов А. С. – М.: ИИнтелЛ, 2016. – 314 с. – (Сер.: Философия творчества).

Материалы Совета были изданы в ряде других издательств. Например, «Глобальное будущее. Конвергентные технологии (НБИКС) и трансгуманистическая эволюция. Под ред. Д. И. Дубровского. М.: Издательство МБА, 2013. – 272 с.», «Дубровский Д. И. Проблема «Сознание и мозг»: Теоретическое решение. М.: Канон+, 2015. – 208 с.» и др. В различных издательствах, возможно, имеются лучшие

полиграфические условия. Однако возникают трудности юридического характера, препятствующие, например, представлению материалов авторов в открытом доступе, в библиотеке РИНЦ, на сайтах заинтересованных организаций.

Учитывая высокую актуальность философских вопросов искусственного интеллекта, первостепенную роль философов в их разработке, а также необходимость систематизации печатных, электронных, аудио- и видеоматериалов, Научный совет РАН по методологии искусственного интеллекта и когнитивных исследований в 2015 году открыл новую издательскую серию «Философия искусственного интеллекта» и поручил собственному издательству «ИИнтелл» провести необходимые для этих целей редакционно-издательские, типографские, полиграфические работы, а также работы по веб-программированию.

Первой книгой серии явилась брошюра «10 лет Научному совету РАН по методологии искусственного интеллекта» (Материалы симпозиума, 26 марта 2015 г., Институт философии РАН, г. Москва / Ред.: Алексеев А. Ю., Дубровский Д. И., Лекторский В. А. – М.: ИИнтелл, 2016. – 142 с.). В брошюре представлены доклады крупнейших отечественных исследователей междисциплинарных проблем искусственного интеллекта.

Настоящий Сборник трудов конференции является второй книгой серии «Философия искусственного интеллекта». Эта книга знаменует начало важного этапа издательской деятельности НСМИИ РАН.

*Главный редактор издательской серии
«Философия искусственного интеллекта»*

В. А. Лекторский

Заместители главного редактора

***А. Ю. Алексеев
Д. И. Дубровский***

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА – 2016

Научные труды
Всероссийской междисциплинарной конференции,
посвященной шестидесятилетию исследований
искусственного интеллекта, 17-18 марта 2016 г.,
философский факультет МГУ имени М. В. Ломоносова,
г. Москва

Под редакцией В. А. Лекторского, Д. И. Дубровского,
А. Ю. Алексеева

Утверждено к печати на 84-ом заседании Научного
совета РАН по методологии искусственного интеллекта
и когнитивных исследований, 13 апреля 2017 г.,
Институт философии РАН, г. Москва

Редактор-корректор: К. К. Кузичева
Составитель: Е. А. Янковская
Переводчик (китайский язык): А. А. Коблякова
Дизайн обложки: Тодор Пожарев
Технический корректор: А. В. Савельев

Издательство «ИИнтелЛ»
109518, г. Москва, ул. Грайвороновская, д. 20 – 171.
book@aintell.info, <http://www.aintell.info>,
+7(499)1771031, +7(926)6201889

Подписано в печать с оригинал-макета 15 мая 2017 г.
Формат 60x84 1/16. Ризография. Гарнитура PT Sans.
Усл.-печ. л. – 21,0. Уч.-изд. л. – 30. Тираж 500 экз.
Заказ № 01 от 15.05.2017 г.

Отпечатано в типографии издательства «ИИнтелЛ»