

Алексеев

Андрей Юрьевич

кандидат философских наук,
координатор научных программ
Научного совета РАН по методологии
искусственного интеллекта,
доцент кафедры информатизации культуры
Московского государственного университета
культуры и искусств

В 1987 г. закончил Ленинградское высшее военное инженерное училище связи им. Ленсовета по специальности программист-математик. В 1987-1996 гг. — научный сотрудник 27 ЦНИИ Министерства обороны СССР (РФ): разработка адаптивных информационных систем. В 1999 г. с золотой медалью закончил философский факультет МГУ им. М.В. Ломоносова по специальности «преподаватель философии», в 2003 г. — аспирантуру этого факультета. В 2004 г. защитил кандидатскую диссертацию на тему «Компьютерное моделирование смысла. Философско-антропологический анализ». В настоящее время работает над докторской диссертацией «Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга».

С 2000 г. — преподавательская деятельность в московских вузах. В последние годы читает курсы «Логика», «Философия», «Информатика», «Информационные системы и технологии», «Проектирование компьютерных систем», «Интеллектуальные информационные системы», «Философия искусственного интеллекта», «Социокультурные проблемы искусственного интеллекта», «Проблема сознания в электронной культуре», «Психосемиотические аналитические системы».

В 2003 г. предложил идею построения «Философии искусственного интеллекта» как прикладной философской дисциплины. Инициировал и участвовал в организации ряда научных мероприятий и научных структур: студенческой (2004 г.) и всероссийской (2005 г.) конференций «Философия искусственного интеллекта», всероссийской конференции студентов, аспирантов, молодых учёных «Искусственный интеллект: философия, методология, инновации», постоянно действующего научно-теоретического семинара «Философско-методологические проблемы искусственного интеллекта» и др.

С 2005 г. участвует в организации и работе Научного совета РАН по методологии искусственного интеллекта: 2005-2008 гг. — ученый секретарь НСМИИ РАН, с 2008 г. по настоящее время — координатор научных программ.

Автор свыше 100 научных работ по тематике электронной культуры, методологии когнитивной науки, философии искусственного интеллекта, проблемы сознания, философских зомби, нейрофилософии, моделирования «знаний» и «смыслов», проектированию искусственной личности и искусственного общества, программированию интеллектуальных систем.



Алексеев А.Ю. КОМПЛЕКСНЫЙ ТЕСТ ТЬЮРИНГА
философско-методологические и социокультурные аспекты

А.Ю.АЛЕКСЕЕВ



КОМПЛЕКСНЫЙ ТЕСТ ТЬЮРИНГА

ФИЛОСОФСКО-МЕТОДОЛОГИЧЕСКИЕ И
СОЦИОКУЛЬТУРНЫЕ АСПЕКТЫ

МОСКВА 2013

Научный совет РАН по методологии
искусственного интеллекта
Московский государственный университет
культуры и искусств

А.Ю. Алексеев

КОМПЛЕКСНЫЙ ТЕСТ ТЬЮРИНГА

ФИЛОСОФСКО-МЕТОДОЛОГИЧЕСКИЕ
И СОЦИОКУЛЬТУРНЫЕ АСПЕКТЫ



**Москва
2013**

УДК 100.32
ББК 32.813
А 471

Рецензенты:

Доктор филос. наук, проф. *О.Е. Баксанский*
доктор филос. наук, доктор техн. наук, проф. *В.А. Глазунов*
доктор филос. наук, проф. *С.В. Лещёв*

Алексеев А.Ю.

А 471 Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. – М.: ИИнтелЛ, 2013. – 304 с.

ISBN 978-5-98956-007-3

Впервые в отечественной и зарубежной философии искусственного интеллекта раскрывается идея комплексного теста Тьюринга. Комплексный тест систематизирует многообразие версий и интерпретаций оригинального теста Тьюринга, играет роль концептуально-категориального инструментария в междисциплинарных исследованиях искусственного интеллекта и выполняет функции дефиниции когнитивных феноменов, конструирования принципиальных проектов интеллектуальных компьютеров, а также рационализации дискуссий о возможностях построения и последствиях применения когнитивно-компьютерных систем.

Книга адресована студентам, аспирантам, междисциплинарным специалистам в области философии, теории и инженерии искусственного интеллекта.

УДК 100.32
ББК 32.813
А 471

ISBN: 978-5-98956-007-3

© А.Ю. Алексеев, 2011 г.
© НСМ ИИ РАН, 2011 г.
© ИИнтелЛ, 2013 г.

ОГЛАВЛЕНИЕ

Предисловие	4
Введение	15
Глава 1. Философские аспекты оригинального теста Тьюринга	17
1.1. От машины Тьюринга к тесту Тьюринга: истоки конструирующей функции комплексного теста Тьюринга.....	17
1.2. Игра в имитацию интеллектуального поведения: истоки дефинитной функции комплексного теста Тьюринга.....	48
1.3. Полемический стандарт теста Тьюринга: истоки критической функции комплексного теста Тьюринга.....	59
Глава 2. Частные тесты Тьюринга	71
2.1. Становление частных тестов Тьюринга	72
2.2. Крупные версии теста Тьюринга.....	77
2.3. Возможности прохождения тестов Тьюринга.....	85
Глава 3. Идея комплексного теста Тьюринга.....	95
3.1. Версии и интерпретации теста Тьюринга.....	95
3.2. Построение комплексного теста Тьюринга	101
3.3. Формулирование комплексного теста Тьюринга.....	105
Глава 4. Дефинитная функция комплексного теста Тьюринга.....	116
4.1. Дефиниция социальных феноменов: «другой» в электронной культуре.....	121
4.2. Дефиниция персональных феноменов: «личность» в электронной культуре.....	128
4.3. Дефиниция ментальных феноменов: «сознание» в электронной культуре.....	148
Глава 5. Критическая функция комплексного теста Тьюринга	162
5.1. Критика технологии: комплексный тест Тьюринга против комплексной технологии динамических экспертных систем реального времени.....	162
5.2. Критика методологии: комплексный тест Тьюринга против комплексного проекта искусственной личности	181
5.3. Критика мировоззрения: комплексный тест Тьюринга против комплексного тестирования в вузовском образовании.....	199
Глава 6. Конструирующая функция комплексного теста Тьюринга.....	222
6.1. Конструирование философской методологии: обоснование стратегической роли проблемы «творчество – компьютер»	222
6.2. Конструирование научной методологии: нейрокompьютинг – базовая технология искусственного интеллекта.....	240
6.3. Конструирование инженерной методологии: комплексная машина Корсакова – Тьюринга – фундаментальный проект интеллектуального компьютера	258
Заключение.....	272
Библиография.....	277

ПРЕДИСЛОВИЕ

Современный человек, неожиданно «заброшенный» в фундированный компьютерными технологиями мир – мир электронной культуры, сталкивается с многочисленными философскими проблемами: технологического конструирования искусственного бытия (В.Л. Макаров), онтологического обживания его адекватными мировоззренческими универсалиями (В.С. Стёпин, В.В. Миронов), этического обустройства (А.А. Гусейнов, Б.М. Шалютин, В.О. Лобовиков, А.А. Скворцов) и эстетического выражения в нём «собственно человеческого» (В.И. Самохвалова). Возникает необходимость пересмотра фундаментальных отношений к смешанной, естественно-искусственной, реальности: её эпистемологического удостоверения (В.А. Лекторский), конструирования и манипулирования её альтерсубъективной (Д.И. Дубровский) и интерсубъективной формами (Н.М. Смирнова). Бурное развитие электронной технологии искусственного интеллекта выпукло обозначает проблемы гносеологической идентичности естественного человеческого мышления (И.Ю. Алексеева, А.Ф. Зотов, А.Н. Кочергин, А.Н. Портнов), философско-методологической компетентности его исследования (Б.В. Бирюков, Т.Б. Кудряшова, А.П. Огурцов, А.И. Уёмов), экстраполяции искусственных форм из биологического наследия (И.П. Меркулов). Развитие электронной культуры актуализирует логические проблемы рационализации общественной жизни (В.К. Финн), семиотические проблемы формирования дискурсов новых форм культуры в её онтологическом, эпистемологическом и системологическом аспектах (В.М. Розин, Ю.М. Резник), герменевтические проблемы понимания её феноменов (В.Г. Кузнецов, Е.Н. Шульга), педагогические (С.П. Капица) и когнитивные (О.Е. Баксанский) проблемы реформирования системы образования, жизненно важные проблемы оптимизации медицины (В.И. Моисеев), аксиологические проблемы обретения нравственной устойчивости в информационном обществе (К.К. Колин) и интеллектуальной безопасности в условиях изощённых информационных войн (И.И. Быстров), инфомедийной идеологической борьбы с религиозным фундаментализмом, сектантством, паранаукой (Э.П. Кругляков, А.А. Ивин). Бесконечный ряд других философских проблем тематизирует рефлексию перехода от традиционной культуры к электронной культуре в сквозном процессе глобальной интеллектуа-

лизации информационного общества (Э.А. Азроянц, А.Н. Чумаков), от личностных гуманистических установок (В.А. Кувакин, Ю.Ю. Чёрный) до общечеловеческих ноосферных регулятивов (Г.С. Смирнов, Н.А. Сляднева, Д.Г. Смирнов).

В размышлении над подобными проблемами акцентируется роль основательной, философской рефлексии по поводу исследований искусственного интеллекта (ИИ). Востребованность в ней подчёркивается традуктивными суждениями: 1) насколько в традиционной культуре важна роль естественного интеллекта как целостной способности упорядочивать чувственные данные, рационально рассуждать и разумно регулировать индивидуальное и социальное поведение, настолько в электронной культуре значима роль интеллектуальных компьютерных систем, реализующих подобные функции; 2) насколько человеческий разум обуславливает расцвет традиционной культуры, настолько креативно-эвристический потенциал интеллектуальных технологий задаёт степень развития электронной культуры.

Актуальность философских исследований ИИ обусловлена конвергентным развитием перспективной материально-технологической инфраструктуры электронной культуры – НБИКС-комплекса (М.В. Ковальчук, Д.И. Дубровский, В.Л. Макаров). Способы функционирования нано-, био-, инфо-, когни- и социотехнологии – принципиально редуцируемы к (квази)алгоритмическим формам электронных преобразований. Компоненты НБИКС-комплекса наукоёмки, методология их интеграции предполагает сложнейшую меж-, кросс- и трансдисциплинарную деятельность (В.И. Аршинов). Эффективности этой деятельности способствуют интеллектуальные компьютерные технологии, которые создают условия организации плодотворного научного творчества, обеспечивают постановку и решение слабо формализованных задач (Д.А. Поспелов, Г.С.Поспелов), а также поддерживают разветвлённый трафик научных коммуникаций (Т.В. Кузнецова), сложный семантико-прагматический поиск информации (Я.Л. Шрайберг), многомерный анализ глобальных информационных ресурсов (О.Б.Сладкова, Н.В.Лопатина, С.А. Чазова) и замысловатые схемы распределения информационных продуктов (А.Б. Антопольский, В.Н. Белоозеров, А.С. Белоцерковский). В целом, реализуя многообразие спектра интеллектуальных функций (Г.С. Осипов), компьютерные технологии стали

неотъемлемы от инфраструктуры современных естественных и социогуманитарных наук. Для синтеза этих наук философия призвана решать традиционные задачи методологической интеграции и мировоззренческой ориентации, но непременно с учётом контекста технологических новаций.

Важность философских исследований искусственного интеллекта обусловлена не столько макрометодологическими интенциями конвергентного синтеза мега-нано-технологий, сколько менее масштабными проектами электронной культуры, современный этап развития которой характеризуется бурным становлением электронных институций (А.Б. Антопольский). В новых социальных кластерах тесно переплетаются элементы традиционной и электронной культуры. Спонтанно возникают и незаметно начинают доминировать в общественной жизни э-музеи, э-библиотеки, э-образование, э-наука, э-искусство, э-государство, э-правительство, э-политика, э-бизнес, э-туризм, э-сообщества, э-семья, э-дом¹ и пр. Становление э-институций непосредственно обуславливается сложными базисными культурными сценариями (В.М.Розин). Философский поход соразмеряет общественные интенции с возможностями компьютерных технологий, осуществляет конструктивно-критическую экспертизу проектов э-культуры, обосновывает социокультурные программы формирования э-институтов, э-хозяйства, э-экономики, э-власти, э-общества, э-личности.

В наибольшей степени значимость философского осмысления искусственного интеллекта как феномена современной культуры (Ю.Ю. Петрунин) и, в частности, как феномена постнеклассической рациональности (В.С. Стёпин) обусловлена обыденной жизнью. В нашей повседневности укоренены интеллектуальные технологии – от бытовых приборов, работающих в недоопределённой динамике внешней и внутренней среды, до глобальных социальных веб-сетей, организация которых предполагает селективную коммуникацию в соответствии со смыслами деятельности виртуальных акторов. В среде электронных технологий и коммуникаций по-иному воспринимаются идеа-

¹ Префикс «э-» обозначает диффузию электронных технологий в социальные практики.

лы социального бытия, трансцендентная проблематика, смысл жизни. Статические архетипы традиционной культуры сменяются канонами программируемой рациональности и, чем сложнее универсалии культуры как программы надбиологической жизнедеятельности (В.С. Стёпин), тем богаче и ярче проявляются феномены электронной культуры. Философская рефлексия по проблемам ИИ стала непосредственно соотноситься с проблемой интерсубъективности [Смирнова, 2011; Дубровский, 2011-1].

**Философские исследования искусственного интеллекта –
это новейшая, когнитивно-компьютерная версия
социальной, персоналистической и экзистенциальной философии**

Существующие варианты философского изучения ИИ не отвечают масштабу этих актуальных задач. Ряд версий вписывает частные методологические концепции ИИ в несоизмеримо широкие теории, например, в общефилософскую онтологию и гносеологию (А. Сломан), категориально-методологический аппарат диалектического материализма (С.М. Шалютин, Г.Н. Поваров, А.К. Зыков, Л.С. Восков). Другие варианты, напротив, слишком узки для социокультурных перспектив, например: программистский (Дж. Маккарти), когнитивистский (М. Боден), нейрокомпьютеринговый (П. Чёрчленд, Ю.Ю. Петрунин, А.В. Савельев, М.А.Рязанов), социально-эпистемологический (М.Ю.Опёнков).

В связи с размытостью методологических оснований ИИ и несоответствием их тенденциям развития электронной культуры автор в начале 2000-х годов предложил ввести в дискурс отечественной науки термин «философия искусственного интеллекта», которому сопутствовали бы логические, методологические, эпистемологические и другие подходы к искусственному интеллекту или аспекты философствования по поводу ИИ. Эти подходы усиленно прорабатывались на протяжении полувека. В отечественную философскую науку предлагается ввести самостоятельное направление исследований. Автор пользуется калькой с англ. «The Philosophy of Artificial Intellegence», которую следовало бы перевести как «размышления по поводу ИИ»¹. Однако, если

¹ В самом деле, возможное мнение, что «философия ИИ» – это о том, как система ИИ начинает философствовать (замечание высказал А.Ф. Зотов).

в таком узком смысле переводить английский термин, тогда нивелируется собственная специфика данной прикладной философии. Острое желание построить самостоятельную философскую дисциплину, предназначенную для выражения программистом своих методологических и мировоззренческих взглядов (дисциплины, подобной «философии медицины», «философии жизни», «философии информации» и пр.), приводит автора к гипотезе, которая в отечественном философском сообществе вначале воспринимается резко отрицательно, в основном, из-за недоумений по поводу вышеупомянутого термина:

**Философия искусственного интеллекта –
относительно автономная сфера междисциплинарных
исследований, определённая основным вопросом
«Может ли компьютер мыслить?»**

Тем не менее, на данную гипотезу обратили внимание видные учёные – Д.И. Дубровский, Н.М. Смирнова, А.Ф. Зотов, А.В. Иванов – в ходе проведения Первых Грязновских чтений (философский факультет МГУ им. М.В. Ломоносова, 2003 г.). В последующем обсуждении проекта «Философия искусственного интеллекта» приняли участие В.К.Финн, В.А. Лекторский, К.К.Колин, И.И.Быстров, А.С.Нариньяни, В.И. Самохвалова, В.В. Миронов, А.А.Гусейнов, В.Г.Кунецов, А.А.Скворцов, О.Е. Баксанский, В.А. Кувакин, Ф.М. Гиренок, Д.Г. Лахути, Г.П. Путилов, Д.В. Быков, А.П. Огурцов, В.И.Аршинов, И.П. Меркулов, А.С. Карпенко, И.К. Лисеев, Н.С. Юлина, В.А. Глазунов, Б.В. Бирюков, А.Д. Королёв, В.В. Трушков, Е.В. Демидова, Л.А. Филлипенко, Н.Х. Сатдинова и др. Результатом обсуждений явилось проведение в 2004 г. конференции «Философия искусственного интеллекта» для студентов и аспирантов института электроники и математики (МИЭМ), председатель – Г.П. Путилов. Конференцию поддержали Д.И. Дубровский, К.К. Колин, А.С. Нариньяни. Она имела большой успех среди молодёжи [Алексеев, 2006-2].

«Взрослая» конференция «Философия искусственного интеллекта» состоялась 18–20 января 2005 г. там же. Сопредседатели: В.Л. Макаров, В.А. Лекторский, Д.В. Быков, Д.И. Дубровский, Г.С. Осипов; уч.секретарь: А.Ю. Алексеев. Конференция вызвала широкий общественный резонанс и оказалась крайне востребованной в условиях бурного становления электронной культуры. Как считает видный философ

и историк кибернетики Б.В. Бирюков, научное мероприятие по проблематике оснований искусственного интеллекта, столь высокого уровня и широкого масштаба, не проводилось с 1965 г. С дискуссий, проходивших на данной конференции, наметился современный этап систематического развития отечественной философии ИИ. Решением конференции явилось создание в марте 2005 г. Научного Совета РАН по методологии искусственного интеллекта при Отделении общественных наук РАН (НСМИИ РАН), председатель – акад. В.Л. Макаров, сопредседатели – акад. В.А. Лекторский, проф. Д.И. Дубровский, ученый секретарь – А.Ю. Алексеев. Немалое значение для широкого развития междисциплинарных исследований ИИ имеет проект всероссийских молодежных конференций «Искусственный интеллект: философия, методология, инновации», предложенный автором в 2005 г. В обсуждении проекта приняли участие Д.И. Дубровский, В.Л. Макаров, К.К. Колин, А.С. Нариньяни, Е.В. Серёдкина, А.С. Сигов, А.Р. Бахтизин, А.И. Бугаков, Е.А. Никитина и др. Конференции призваны закрепить среди молодёжи концепцию философии ИИ как нечто само собой разумеющееся. Первые две конференции состоялись в Москве (2006 г.) и в Санкт-Петербурге (2007 г.). На первой конференции была создана молодёжная секция НСМИИ РАН. Ключевую роль в этом сыграла Е.В. Серёдкина – председатель секции в 2006 – 2007 гг. Однако идея автора о проведении ежегодных молодежных конференций на базе крупных вузов страны, к сожалению, не получила развитие. Последующие конференции проводятся только в Москве на базе МИРЭА (ученый секретарь – Е.А. Никитина). Такой подход не способствует масштабному расширению исследований.

Тем не менее, активность изучения методологии искусственного интеллекта в рамках мероприятий НСМИИ РАН, в научной и технической организации которых автор принимает непосредственное участие, убедительно подтверждает тот факт, что тематика философии ИИ постоянно расширяется и конкретизируется, а количество исследователей растет в дисциплинарном, региональном и возрастном масштабах. Искусственный интеллект стал существенным атрибутом современной – электронной – культуры. Он достоин самостоятельного направления в философской науке. Если так, то необходим категориально-концептуальный аппарат междисциплинарных исследований ИИ, с которым могут работать многопрофильные специалисты. Применительно

к рефлексии над проблематикой когнитивно-компьютерных технологий этот методологический инструментарий должен играть роль, аналогичную, скажем, категориям общей философии.

Эталоном междисциплинарных дискуссий по изучению базовых принципов создания, функционирования и применения в обществе компьютеров, которым приписывается (атрибутируется) «интеллект», принято считать *тест Тьюринга*. Показателем его международного междисциплинарного признания является большое разнообразие различных версий, модификаций и интерпретаций, предложенных на протяжении шестидесяти лет осмысления его роли для философии, науки, техники, искусства и культуры. Тесты Тьюринга атрибутируют компьютерным системам разнообразный спектр витальных, психических, персональных, социальных феноменов, внося существенный вклад в их изучение. Однако отдельное и разрозненное применение частных тестов не обеспечивает многостороннее и многомерное изучение когнитивных феноменов в контексте гиперсложного пространства электронной культуры. Необходимо объединить тесты в комплексном формате.

В данной книге отстаивается следующее положение:

Комплексный тест Тьюринга является категориально-концептуальным инструментарием междисциплинарных исследований искусственного интеллекта

Комплексный тест обеспечивает идентификацию, спецификацию, систематизацию, дифференциацию, унификацию, координацию и интеграцию сложных знаний, методов и средств, касающихся проблематики искусственного интеллекта.

Основными теоретическими источниками являются многочисленные работы, посвящённые тесту Тьюринга, включая оригинальную статью «Вычислительные машины и интеллект» (1950 г.), в которой сформулированы категориальные идеи атрибуции «мышления» машине, принципиальный проект интеллектуального компьютера и типовой полемический стандарт междисциплинарных дискуссий по поводу интеллектуальных машин. Со многими частными тестами русскоязычные исследователи знакомятся впервые.

Особая роль приписывается статье отечественного пионера искусственного интеллекта С.Н. Корсакова «Очерк о новом способе ис-

следования посредством машин для сравнения идей» (1832 г.), впервые в полном объеме переведённую спустя 175 лет с французского языка на русский под научной редакцией автора. Для раскрытия принципиальных возможностей реформирования современного компьютеринга на основе революционных идей С.Н. Корсакова использовались классические теоретико-алгоритмические модели, что выявило двойственность проекта Корсакова относительно традиционных принципов вычисления и его своеобразную постнеклассическую специфику. В отличие от машины Тьюринга как метафоры работы «математика», пребывающего в «рефлексии» автоматных состояний, обзорающего и преобразующего «мир» посредством узкого канала символьной репрезентации, в машине Корсакова действует художник, «энергично пробивающий» многочисленные «окна в реальность» для нерефлексивного её «схватывания».

Текст книги сложился в ходе дискуссий в рамках постоянно действующего ежемесячного научно-теоретического семинара НСМ ИИ РАН «Философско-методологические проблемы искусственного интеллекта». На юбилейном, пятидесятом заседании семинара, которое состоялось 16 декабря 2011 г., автору посчастливилось подытожить результаты изучения комплексного теста Тьюринга¹. Широкий спектр когнитивных феноменов, который охватывается комплексным тестом и многообразие дискуссий, сопровождающих концепции построения интеллектуальных компьютеров, послужили иллюстрацией к теме основного доклада Ю.Ю. Петрунина «Искусственный интеллект как феномен современной культуры».

Эта книга полезна тем, что представленные в ней выводы и теоретические положения могут быть использованы в качестве дополнения существующих и разработки новых учебных курсов для студентов высших учебных заведений и курсов «история и методология науки» для аспирантов. Для развития образовательной стратегии исследований ИИ важно включение результатов работы в учебные курсы по специальности «Интеллектуальные системы в гуманитарной сфере», которую

¹ С первым докладом на тему «Роль Теста Тьюринга в методологии искусственного интеллекта» автор выступил на семинаре НСМ ИИ РАН 26 апр. 2006 г.

впервые апробировали Е.М. Бениаминов, Д.Г. Лахути, В.К. Финн и другие учёные РГГУ.

Критический потенциал комплексного теста Тьюринга может быть использован в экспертизе государственных и общественных программ развития электронной культуры России. Это – очень экономная экспертиза. Она не требует сложных социологических исследований и дорогостоящих натурных экспериментов с внедрением компьютеров в экономику, политику, образование, здравоохранение, оборону. Требуется «сесть в кресло» и обдумать проблемы электронной культуры в контексте нарабатанных мысленных экспериментов, яростных концептуальных баталий, убедительных логических аргументов за и против возможностей интеллектуальных и неинтеллектуальных компьютеров.

Рефлексия по поводу комплексного теста Тьюринга очень эффективна и, самое главное, отвлекает от постылой унылости бытия квазикапиталистического общества потребления.

Автор благодарен инженерам-математикам курсантской группы Ленинградского высшего военного инженерного училища связи им. Ленсовета – В. Боровкову, А. Денисову, В. Ткачуку, В. Кузнецову, Р. Мельничуку, И. Корневскому, Д. Гуль, А. Клушу, А. Зяблову, В. Неугасову и другим – за конструктивный контекст первых «школьных» шагов автора в изучении ИИ (1982 – 1987 гг.).

Особая благодарность военным начальникам из 27-го ЦНИИ МО – И.И. Быстрову, В.В. Дееву, В.С. Кавтыреву, Ю.Н. Тронину – за вовлечение автора в серьёзные работы по проектированию сложных интеллектуальных систем военного назначения, а также друзьям-офицерам, программировавшим подобные системы, – А. Кулянице, А. Плюснину, В. Антонову, С. Филлипенко, А. Вертоградову, С. Кирилову (1987 – 1996 гг.).

Автор также благодарит замечательных преподавателей философского факультета МГУ им. М.В. Ломоносова – П.В. Алексеева, А.Ф. Зотова, А.И. Болдырева, В.А. Бочарова, Е.В. Брызгалину, В.В. Васильева, А.Г. Гаджикурбанова, Ф.И. Гиренок, А.И. Иванова, В.В. Миронова, В.Г. Кузнецова, В.Н. Кузнецова, Е.В. Косилову, Ю.Б. Сенчихину и многих других. Особая благодарность –

В.А. Кувакину, который, помимо неоскептического мировоззрения, привил навыки организации научных мероприятий и научных сообществ.

Глубочайшая благодарность – Н.М. Смирновой. Без её научного руководства не состоялись бы ни мой диплом на философском факультете МГУ, ни кандидатская диссертация. Все научные мероприятия, которые автор организовывал на всероссийском уровне (их несколько десятков), вначале получали её одобрение.

Автор бесконечно благодарен Д.И. Дубровскому, своему главному учителю. Опираясь на его авторитет, автор смог отстаивать положения аналитической философии искусственного интеллекта, которая совсем недавно не признавалась в кругах вузовской и академической науки. Неценимой так же была его моральная и дружеская поддержка.

В.И. Самохваловой автор благодарен за то, что она впервые в широкой печати отметила его статью (проект искусственной личности, 1998 г.) и с тех пор конструктивно критикует и продвигает его работы.

Особая признательность – К.К. Колину и Т.Б. Кудряшовой, которые способствовали оформлению идеи комплексного теста Тьюринга в данной монографии и в докторской диссертации.

Андраник Тумасян, Елена Серёдкина (организатор молодежной секции НСММИ РАН), Андрей Королёв, А.Н.Виноградов, Наташа Трушкина (Клюева), Татьяна Кураева (Конькова), Алла Дрозд, Сабир Гейбатов, Сергей Лещёв, Сергей Игнатов, Маргарита Пилюгина, Юлия Моркина, Дарья Полунина и др. очень помогали в содержательной и организационной работе.

Благодарен за плодотворное обсуждение ряда спорных положений – В.А. Лекторскому, В.И. Аршинову, О.Е. Баксанскому, В.Л. Васюкову, В.А. Глазунову, А.С. Карпенко, Е.Н. Князевой, Т.Б. Кудряшовой, Д.Г. Лахути, И.П. Меркулову, А.П. Огурцову, С.А. Павлову, Ю.Ю. Петрунину, В.М. Розину, А.В. Савельеву, В.К. Финну, Е.Н. Шульге, Н.С. Юлиной.

В.Л. Макаров, А.Р. Бахтизин помогли автору оценить главное назначение комплексного теста Тьюринга: для методологии искус-

ственных обществ. Этот новейший проект интегрирует существующие и перспективные направления искусственного интеллекта.

Особенно благодарен пионерам отечественной философии ИИ – студентам МИЭМ¹, работы которых вошли в Сборник студенческой конференции «Философия искусственного интеллекта», 20 апреля 2004 г. [Алексеев, 2006-2,4]. Пользуясь привилегиями преподавателя социальной философии (2003 – 2005 гг.), автору удалось воодушевить ребят на переводы статей с англ. и немец. языков и написание собственных эссе, рефератов, докладов, посвященных тесту Тьюринга и другим методологическим аспектам ИИ. Во многом благодаря удачно проведенной студенческой конференции 2004 г. состоялась «взрослая» конференция с таким же названием в 2005 г.

Автор благодарен за поддержку исследований электронной культуры со стороны коллег по Московскому государственному университету культуры и искусств – А.Б. Антопольскому, Т.В. Кузнецовой, О.Б. Сладковой, Н.А. Слядневой, А.С. Белоцерковскому, В.Н. Белоозерову, А.А. Гук, О.В. Шлыковой, Д.Г. Делицину, О.В. Громовой, А.М. Варэлису, С.А. Чазовой, Н.В. Лопатиной и др.

Очень благодарен сестре *Светлане*, племянницам *Насте* и *Наташе*, племяннику *Никите* и его супруге *Ольге*, сыновьям – *Даниле* и *Василию* – за терпение, понимание и посильную помощь. Без поддержки моей любимой жены и коллеги по философскому цеху – *Екатерины* – данная работа не имела бы смысла.

И, самое главное, автор фундаментально обязан своим дорогим родителям – *Юрию Ивановичу* и *Нине Александровне Алексеевым*, без которых вообще ничего бы не было.

¹ К сожалению, Московского института электроники и математики как самостоятельной вузовской единицы сегодня не существует – он вошел в состав НИУ ВШЭ.

ВВЕДЕНИЕ

Идея теста Тьюринга инспирирована проблемой интерсубъективности. В контексте э-культуры данную проблему характеризует пример, который автор назвал «Девушка по переписке» [Алексеев, 2006-1].

Молодой человек ищет спутницу жизни и обращается к интернет-службе знакомств. Задаёт параметры запроса – возраст, рост, объёмы, цвет волос, знак зодиака, привычки и пр. В ответ – ряд кандидаток. Начинает переписываться. Особенно нравится девушка Элиза. Переписка длится более полугода – в выборе нельзя ошибиться! Беседы ведутся на самые разные темы: семьи, детей, секса, быта, отдыха, зарплаты, экономики, политики, литературы, китайской кухни и пр. Играют в шахматы с переменным успехом. Девушка поражает объемом знаний, лёгкостью в общении, изысканностью стиля. И глаза у неё выразительные. Наконец молодой человек не выдерживает и назначает встречу на площади у памятника. Приезжает с букетом роз. Ждёт. Час, два, три. Никто не приходит. Вернувшись ни с чем, пишет гневное письмо. В ответ – «Прости, милый! Я – Электронная Лиза, компьютерная программа»¹.

Молодому человеку не повезло – и поделом ему. Во-первых, он не потрудился укрепить адекватность своих представлений действительному положению дел, не задумался об интерсубъективной проблематике в контексте компьютерной технологии. Во-вторых, он слишком доверился наивной феноменологии восприятия *другого*, атрибутировав телекоммуникационному собеседнику параметры человека. В-третьих, он не приложил усилий эпистемологического удостоверения общения с человеческой личностью, а не с программой, реализующей функции философских зомби.

¹ На выбор имени «девушки» повлияло название широко известной программы «понимания» естественного языка, – ELIZA, разработанной Дж. Вейзенбаумом в 1966 г. [Weizenbaum, 1974; 1976; 1996]. Программа состояла всего из двух сотен операторов на языке Бейсик и незатейливого алгоритма «понимания» задаваемых ей вопросов. При этом ELIZA настолько умело имитировала речевые способности собеседника, что секретарша Вайзенбаума при отсутствии шефа очень любила с нею «поболтать». В основе концепции программы лежала психотерапевтическая техника диалога с пациентом. Сегодня программы, у которых в основе диалога лежит такая же модель, принято называть «ELIZ'-а-подобные программы» [Weizenbaum, 1996].

В шутку замечено, что в Интернете выдавать себя за человека может не только программа, но и животное, например, натренированная собака. Питер Стейнер в знаменитой карикатуре «В Интернете никто не знает, что ты – собака!» (1993 г.) [New Yorker, 1993] изобразил пса, сидящего перед компьютером и нажимающего на клавиши. «И в самом деле!» – восклицает В. Рапопорт по поводу этой карикатуры [Rapoport, 2000, P. 469]: «В Интернете никто не знает, кто с кем на самом деле общается. Это неведение может привести к самым серьёзным последствиям в реальной жизни. Если я ввожу номер кредитной карточки через Web-сайт для покупки книги, то неизвестно, куплю ли я эту книгу или выдам свой секретный код мошеннику. Компьютерная безопасность определяет способ социального благополучия и защиты личности, и даже жизнь или смерть. Если моя дочь общается посредством чата с лицами противоположного пола примерно её возраста, то кто знает, может быть, она общается с серийным убийцей. Обычно, правда, мы предполагаем такие опасности и полагаем, что общаемся не с мошенниками, убийцами, собаками. Или – как в нашем случае, с компьютерами».

Утверждения В. Рапопорта (2000 г.) сегодня устарели на фоне экспоненциального развития информационно-коммуникационных средств, сопровождаемого, в свою очередь, громкими судебными делами по поводу преступлений, «соучастниками» которых выступают компьютеры. По всей видимости, в недалёком будущем воплотится и сценарий «Девушка по переписке». Тогда актуальнейшей проблемой станет поиск методов и средств распознавания отличий между человеком и компьютерной системой.

Впервые подобный сценарий представил научному сообществу А. М. Тьюринг в небольшой статье «Вычислительные машины и интеллект». Статья была опубликована в 1950 г. в авторитетном журнале «Mind» [Turing, 1950] и стала самой цитируемой в литературе по философии сознания, когнитивной науке, искусственному интеллекту [Akman, Blackburn, P. 391]. Широкие обсуждения и жаркие дискуссии по основному вопросу статьи «Может ли машина мыслить?» породили новейшую философию аналитической традиции – философию искусственного интеллекта. Изучению и развитию сценария А. Тьюринга в контексте электронной культуры посвящена данная работа.

ГЛАВА 1. ФИЛОСОФСКИЕ АСПЕКТЫ ОРИГИНАЛЬНОГО ТЕСТА ТЬЮРИНГА

1.1. От машины Тьюринга к тесту Тьюринга: истоки конструирующей функции комплексного теста Тьюринга

Исследователи теста Тьюринга, изучая его возможности и предлагая собственные версии, как правило, начинают с фиксации ключевой идеи знаменитой статьи Алана Мэтисона Тьюринга «Вычислительные машины и интеллект». Идея состоит в замене неясного вопроса «Могут ли машины мыслить?» на более простой вопрос, который целесообразно применять в контексте изучения «машинного мышления». Сложность исходного вопроса возникает при попытке внятной дефиниции «интеллекта». Это – более сложная проблема, нежели чем, скажем, «самая сложная проблема сознания», позиционированная Н. Блоком и, далее, Д. Чалмерсом при объяснении отношения феноменальное/физикальное. В этих объяснениях мы *как бы* обладаем априорными дефинициями «феноменального», «физикального». Но что такое «интеллект» как фундаментальная внутренняя способность разумного человеческого существа (*homo sapiens*), благодаря которой мы владеем этими дефинициями? Вот самая тайная из тайн. Кстати, для А. Тьюринга не менее трудной оказалась проблема дефиниции «машины». В самом деле, что такое «машина»?

А.Тьюринг поступает по-инженерному прямо и конкретно. Он упрощает проблему, применив простую *антиэссенциалистскую стратегию*, в соответствии с которой надо исключить поиск глубинной сущности понятий. Это – неблагоприятная и бесперспективная работа. Для инженерного успеха в построении интеллектуальных машин не нужны метафизические изыскания. Вопрос о мыслящих машинах следует перевести в практическую плоскость изучения: «Можно ли создать машину (точнее, универсальную цифровую машину), которая будет способна подражать интеллектуальному поведению человека?». Предмет вопроса переносится в контекст *игры в имитацию* (ИВИ) интеллектуального поведения. ИВИ имеет концептуальный статус, т.е. она представлена в формате мысленного эксперимента. Участниками игры вы-

ступают и люди и компьютеры. Если для стороннего наблюдателя (судьи), который не имеет возможности наблюдать игроков, процесс диалогового общения с компьютером не отличим от общения с человеком, то *компьютер может мыслить*. Всё очень просто.

Большинство специалистов в области когнитивной и компьютерной наук полагают, что для философии, теории и практики искусственного интеллекта данная статья явилась основополагающей. Соответственно, и идея теста Тьюринга, представленная в ней, выступает фундаментом для всего последующего строения ИИ. Это так. Однако, на мой взгляд, изучение данной работы, несомненно крайне важной, но взятой отдельно и изолированно от других тьюринговых работ, не позволяет в должной мере осветить идею теста и оценить её роль для становления философии ИИ. Например, при использовании идеи Тьюринга в контексте методологических исследований проблематики сознания неправомерно утверждается, что «тьюринговский тест на интеллект искусно преодолел всё бескрайнее философское болото, возникшее из проблемы отношений разума и тела (mind-body problem)» [French, 1990, Р. 55]. Оригинальный тест в «болото» отношений феноменального/физи-кального даже и не засасывало, поскольку А. Тьюринг не был философом и никогда не причислял себя к их числу. В вязкую тину метафизических словоплетений, из которых состоит современная философия сознания, попали интерпретации теста, но не оригинальный тест. Э.Ходгис, известный биограф А. Тьюринга отмечает следующее: «Алан Тьюринг, наверное, посмеялся бы над тем, чтобы назвать себя не только великим философом, но и просто философом. Он считал себя математиком» [Hodges, 2010]. Тем не менее, версии тьюринговой идеи имеют важные философские приложения. Поэтому совершенно правомерно в 1996 г. тот же Э. Ходгис публикует книгу «Тьюринг» в известной серии «Великие философы» [Hodges, 1997].

Факт того, что статья Тьюринга стала одной из наиболее цитируемых работ в современной философской литературе, объясняется просто. А. Тьюринг перенёс достаточно строгое математическое понятие вычислимости и алгоритма на традиционные проблемы души и тела, свободы воли и детерминизма. Вечные мировоззренческие вопросы получили современную трактовку в эпоху электронной культуры, в алгоритмизированном и компьютеризированном мире. Причем трансформация «вечных идей» в компьютерном и телекоммуникационном

контексте осуществилась не путем высоких теоретических изысканий, но посредством экспликации обыденного процесса общения между людьми. Некоторые из людей (возможно все) на поверку оказываются машинами, играющими роли людей.

Так как А. Тьюринг не был философом, то перед осуществлением философских интерпретаций теста следует обратить внимание на математические работы. Именно в них кристаллизовывалась идея теста на машинный интеллект. И собственно математиком, а не философом был А.М. Тьюринг, по крайней мере он сам в этом признавался. Математическая линия зарождения идеи теста явно прослеживается в концептуальной ветви, произрастающей из знаменитой конструкции, предложенной А. Тьюрингом в 1936 г. [Turing, 1936] и получившей в 1937 г. собственное имя «машина Тьюринга».

Машина Тьюринга – инженерная формализация интуитивного понятия «алгоритм»

Из курса школьной информатики известно, что алгоритм – это фундаментальное понятие математики и программирования. Имеется государственный стандарт на понятие алгоритма, и данную дефиницию не отменила переориентация страны с социалистической формы хозяйствования на капиталистическую. В соответствии с ГОСТ 19.004-80 (он сегодня заменён), «алгоритм — это точное предписание, определяющее вычислительный процесс, ведущий от варьируемых начальных данных к искомому результату». В данном понятии имплицитно фиксируются следующие свойства алгоритма: *детерминированность* – причинная предопределённость действий в каждом возможном случае и однозначность результата работы при заданных исходных данных; *массовость* – возможность варьирования исходных данных в определенных пределах и, соответственно, пригодность алгоритма для решения некоторой совокупности задач определённого класса; *результативность* – для любых допустимых исходных данных через конечное число шагов работа должна завершиться; *дискретность* – разбиение алгоритмического процесса на отдельные элементарные этапы, возможность реализации которых очевидна, а результат выполнения каждого элементарного этапа определён и понятен.

Все эти определения и свойства – интуиции по поводу понятия алгоритма. В теории алгоритмов выделяется несколько подходов к

формализации этого интуитивного понятия. На наш взгляд, классическими подходами представляются математическая, лингвистическая и инженерная трактовки. Другие интерпретации – В.А. Успенского, А.Н. Колмогорова и, скажем, популярная сегодня интерпретация Г.Чайтина, сводимы к классическим подходам. Помимо этого, в исходных трактовках имеются предпосылки для репрезентации знаний в современных компьютерных системах, за исключением, конечно, нейрокompьютинга и, соответственно, коннекционизма.

Математическая трактовка понятия алгоритма связывается с исследованиями Алонзо Чёрча в области теории вычислимых функций. Отделив понятие функции от понятия множества посредством введения операции функциональной абстракции (λ -операции), Чёрч построил логическое исчисление (λ -исчисление) и в 1936 г. выдвинул идею, что понятие алгоритмической вычислимости выражается посредством частично определённых рекурсивных функций [Барендретт, 1985]. На основе λ -исчисления в середине 20-го века был создан первый язык программирования интеллектуальных компьютерных систем Lisp (его реализовал Джон Маккарти [McCarthy, 1960]). Сегодня Lisp имеет мощные приложения в области аппликативного, комбинаторного, функционального и других парадигм программирования, которые ориентированы на построение «чистых» программных инструкций, абстрагированных от среды реализации. Недавно фирма Microsoft, известная инертностью к нормальным инновациям, включила в состав Visual Studio язык функционального программирования F#.

Лингвистическая трактовка понятия алгоритма предложена в конце 1940-х годов советским математиком Андреем Андреевичем Марковым в теории нормальных алгорифмов [Марков, 1984]. Нормальный алгорифм применяется к словам определённого алфавита и задаётся посредством дискретной последовательности элементарных шагов переписывания строк, определяемых формулами подстановки. Данный подход сегодня активно используется в одном из классических способов программирования интеллектуальных систем – в производственном программировании.

Инженерная трактовка понятия «алгоритм» непосредственно связывается с машиной Тьюринга. В состав машины Тьюринга (МТ) входят разделённая на ячейки бесконечная в обе стороны лента (воз-

можно МТ, которые имеют неограниченное множество лент бесконечной длины), а также управляющее устройство, способное находиться в одном из множества состояний. Число возможных состояний управляющего устройства конечно и точно задано. Управляющее устройство работает согласно правилам переходов, которые, собственно, и представляют алгоритм, реализуемый данной машиной Тьюринга. Каждое правило перехода предписывает машине, в зависимости от текущего состояния и наблюдаемого в текущей клетке символа: стереть старый символ, оставить его без изменения, записать в эту клетку новый символ, остаться в предыдущем состоянии, перейти в новое и/или переместиться на клетки влево или вправо. Некоторые состояния машины Тьюринга помечаются как терминальные, и переход в любое из них означает конец работы, остановку алгоритма. Таким образом, понятие алгоритма эксплицируется способом функционирования машины [Turing, 1936]. В работе 1950-го года понятие «интеллекта» эксплицируется почти аналогично – способом функционирования машины, имитирующим интеллектуальное поведение.

В принципе данные три трактовки алгоритма эквивалентны. Факт эквивалентности выражен в знаменитых тезисах, которые доказать невозможно, но они представляются как очевидные положения. Тезис Чёрча – Тьюринга в самой общей форме гласит, что любая интуитивно вычислимая функция является частично вычислимой, или, что равнозначно, может быть вычислена некоторой машиной Тьюринга. В свою очередь, «принцип нормализации» фиксирует, что любой нормальный алгоритм, в смысле А.А. Маркова, эквивалентен некоторой машине Тьюринга и, наоборот, любая машина Тьюринга эквивалентна некоторому нормальному алгоритму. А раз так, то и алгоритм Маркова эквивалентен частично рекурсивной функции Чёрча, и, соответственно, машине Тьюринга.

Несмотря на общепринятое мнение о равнозначности трактовок понятия «алгоритм», машина Тьюринга стоит на вершине *эвристической компетентности*. Во-первых, машину Тьюринга легко вообразить в формате мысленного примера, в отличие от частично-рекурсивных функций и формальных грамматик, где требуется дополнительная конкретизация механизма, выполняющего исчисления и комбинирующего текст. Во-вторых, и это – главное, в машине Тьюринга двuasпектно совмещены два плана – формальный (символы на ленте) и каузальный

(конкретное устройство чтения/записи символов, переходящее из одного состояния в другое).

Поскольку в машине Тьюринга выполняется соотношение *формальное/каузальное*, отсюда недалеко до идеи функционализма – доминирующей парадигмы современной философии сознания. Для философии сознания главным вопросом был и остаётся вопрос отношения «идея/материя», «дух/тело», «психика/мозг», «феноменальное/физическое». Функционализм обозначил ряд новых дихотомий: «функциональное/реальное», «когниция/реализация», «формальное/каузальное». Машинный функционализм породил современный компьютеризм (computationalism), в котором «формальное» – это программная дескрипция, «каузальное» – физический процесс функционирования компьютера. Компьютизм распространился на когнитивные феномены ментального, витального, персонального, социального содержания. Затем недалеко оказалось и до социокультурных феноменов: свободы, бессмертия, смысла жизни, общественного идеала. Как видим, простой замысел инженера получил колоссальный резонанс во всех сферах современного компьютеризованного мира – мира э-культуры.

Функционализм машины Тьюринга – основная «функция» машины Тьюринга в философии искусственного интеллекта

Становление и развитие функционалистской парадигмы сознания и мышления обусловлено развитием кибернетики, теории информации, семиотики, системных и структуральных исследований. Функционализм – одно из главных теоретических событий аналитической философии 20-го века [Дубровский, 1994, С. 21]. Функционализм предлагает особый подход к проблеме «сознание-мозг-компьютер». При решении данной проблемы ищутся ответы на вопросы: что такое психическое, что такое мысль, где мысль существует, что делает мысль мыслью, что делает боль болью? Ответы на эти вопросы связываются с функциональной организацией целостного организма.

Картезианский дуализм утверждает, что ментальное («боль», «мысль», «переживание» и пр.), пребывает в особой духовной субстанции. Бихевиоризм идентифицирует психику с поведенческой функцией. «Боль», «мысль» – это диспозиции поведения. Они не имеют внутреннего, интенционального содержания. Физикализм в своей основной версии отождествляет психическое с мозговыми процессами. «Боль», «мысль» –

это физические процессы, описываемые нейрофизиологическими терминами, референты которых либо соотносятся с народно-психологическими терминами, либо, по утверждению элиминативистов, не соотносятся с ними, но лишь с прогрессивно уточняемыми научными понятиями.

Функционализм утверждает, что психические явления – это, во-первых, отношения между чувственными данными, поведенческими реакциями и психическими состояниями. Эти функциональные отношения выполняют специфические роли в составе психики как целостной системы. Во-вторых, это – функции, реализуемые материальными, в частности мозговыми процессами. «Мысль», «боль» – это функциональные состояния нейромозговых коррелятов или других систем, субстанционально различных, но функционально инвариантных мозгу. В *исторически первой и оригинальной версии функционализма* такие состояния концептуально отождествляются с состояниями машины Тьюринга. Данная версия функционализма получила название «*функционализм машины Тьюринга*». Часто употребляются синонимы: «*машинный функционализм*», «*функционализм состояний машины Тьюринга*», «*функционализм машинных состояний*».

Концепция функционализма машины Тьюринга (ФМТ) имеет давнюю историю и многих последователей. Недавно вышли две крупные историко-философские работы на эту тему: Роберта Харниша «Сознание, мозг, компьютер: историческое введение в фундаментальные основы когнитивных наук» [Harnish, 2001] и Роберта Камминса: «Сознание, мозг, компьютер: основы когнитивных наук. Антология» [Cummins, 2000]. В указанных антологиях наиболее почётное место отводится основателю машинного функционализма – Х. Патнэму, который, кстати, впоследствии радикально отказался от своих «тьюринговых» взглядов на проблематику сознания [Пассмор 2002, С. 106–107].

Идея функционализма машины Тьюринга и её экспликация

Впервые концепция функционализма машины Тьюринга была предложена Х. Патнэмом в небольшой статье «Сознание и машины» [Putnam, 1960]. Сегодня машинный функционализм раскрывается следующим образом (цит. по [Levin, 2004]; см. также [Cummins, 2000]): «Обладающее сознанием существо можно рассматривать как машину Тьюринга (идеальный компьютер с конечным числом состояний), рабо-

та которой специфицируется множеством инструкций машинной таблицы или программы, каждая из которых имеет следующую форму. Если машина находится в состоянии S_i , получает входной символ I_j , то она перейдёт в состояние S_k и вырабатывает выход O_l (для конечного числа состояний, входных и выходных символов)».

В данном суждении чётко обозначается перспектива машинного функционализма – в способе реализации психических явлений, инвариантном относительно их материального носителя. «Значения» функции – феномены психики, сознания, самосознания, мышления и пр. – можно получить при различных «аргументах». В область определения такой «mind/brain»-функции входят, например, электрохимические мозговые процессы. Кстати, первый «функционалистский» проект искусственной личности (при данном понимании «функции», определённой на этих процессах) впервые предложила в 1818 г. М. Шелли в образе демона Франкенштейна [Шелли, 1990]. В область определения «функции» входят не только близкие нам нейральные системы животных и других живых существ, но и воображаемая тина марсиан, космологические системы Соляриса и пр. Наконец, аргументами «функции» являются дискретные электронно-цифровые преобразования логических вентилей – редукционный базис современного компьютеринга. Обширность области определения «психической функции», задание других функций относительно результатов предыдущих реализаций, входов и выходов системы, и, в конечном итоге, определение ролевой структуры психических функций в составе целостной системы (организма) позволяет сформулировать один из главных принципов функционализма – *возможность множественной реализации психического*. По словам Н.Блока, в такой перспективе программа изучения психики как функциональной организации системы исключительно человеческого индивида приобретает либеральный характер и дополняется «реализаторами» самой различной, нечеловеческой и небиологической природы.

Поскольку имеется многообразие способов физической реализации одной и той же машины Тьюринга, то тождественные ментальные феномены могут быть присущи как разным людям, так и, скажем, инопланетянам, у которых нет человеческого мозга, но имеются состояния мозговой тины, эквивалентные логическим состояниям машины Тьюринга [Putnam, 1960]. Через пятнадцать лет в работе «Философия и наша ментальная жизнь» [Putnam, 1975] Х. Патнэм более чётко форму-

лирует «антропологическую» позицию машинного функционализма: 1) человек – это машина Тьюринга; 2) психологические состояния человека – это состояния машины Тьюринга или дизъюнкции состояний машины Тьюринга. В последующем, пытаясь совместить социокультурную проблематику с машинным функционализмом, Х. Патнэм считал концепцию функционализма глубокой ошибкой. В итоге он скатился к солипсизму в решении вопроса множественной реализации тождественных психических состояний на различных физических субстратах: ментальные состояния – это мои личные состояния и они предполагают знание глубоких социальных отношений между людьми, которые невозможно описать ни машинными состояниями, ни дизъюнкциями машинных состояний [Пассмор, 2002, С. 107]. Сознание – сугубо личностный феномен, и ни о какой компьютерной реализации *моего* сознания не может быть и речи.

Несмотря на скептицизм позднего Х. Патнэма, исходные оптимистичные идеи машинного функционализма сегодня активно эксплуатируются в функционалистских теориях сознания, в общем, и компьютеристских теориях, в частности [Horst, 2003].

Рассмотрим две превосходные архетипичные иллюстрации функционализма машины Тьюринга, предложенные Н. Блоком более тридцати лет назад в работах «Проблемы функционализма» [Block, 1978] и «Функционализм» [Block, 1980]. В первой работе предлагаются мысленные эксперименты, которые сегодня стали классическими: «Гомункульная голова» (1978 г.) и «Китайская нация» (1978 г.). Вторая работа имеет номинативную аналогию со знаменитой «Китайской комнатой» Дж. Серля¹. В этих работах достаточно ясно излагается идея машинного функционализма и устанавливается его связь с теоретическим функционализмом (психофункционализмом) посредством методики Рамсея – Льюиса.

Мысленный эксперимент «Гомункульная голова». В статье «Проблемы функционализма» Н. Блок предлагает эксперимент, в кото-

¹ Дж. Серль, кстати, никак не упоминает о «Китайской нации» в работе «Разумы, мозги, программы» [Searl, 1980], в которой предлагает мысленный эксперимент «Китайская комната». В последующем комментаторы «Китайской комнаты» восстановили справедливость, указав, что отмеченная работа Н. Блока является её концептуальной предшественницей [Cole, 2009].

ром участвуют гомункулы – невидимые маленькие «человечки», сопровождающие и реализующие психическую деятельность человека, тем самым несколько отличаясь от фаустовских «гомункулов» – средневековых алхимиков, которые стремились создать искусственных личностей посредством манипуляций со спермой, кровью, навозом и пр. В работе Н. Блока «гомункул» употреблён в смысле У. Пенфилда как элемент функциональной организации моторных и сенсорных зон в коре головного мозга. Однако, на наш взгляд, предпочтительнее было бы употребить термин «демон Сэлфриджа», чем «гомункул Пенфилда»: «пандемониум Сэлфриджа», дуалистически эксплицирующий перцептивные способности [Линдсей Норман, 1974, С.78], очень похож на «гомункулную голову» Н. Блока. В «блоковской голове» (именно так иногда называют гомункулную голову) выражена простая версия машинного функционализма: психические явления описываются таблицей машины Тьюринга. Каждое ментальное состояние отождествляется с некоторым логическим состоянием, которое, в свою очередь, характеризуется структурой и содержанием таблицы, т.е. содержимым ячеек при корреляции их с другими ячейками и с таблицей в целом¹. Вход таблицы, которую предлагает Н. Блок, специфицирует импульсы нейронов органов чувств, выход – импульсы в моторно-двигательных нейронах.

Строение гомункулной головы таково: помимо «пандемониума Сэлфриджа», архитектура её очень схожа с платоновской пещерой. Напомним, что Платон применил образ пещеры для прояснения того, каким образом люди причащаются идее, которая бросает свет истины на стены пещеры. Пещерные люди Платона воспринимают отражения. Так же и гомункулы Н. Блока истине в глаза не смотрят. Им этого и не надо. Главное для них – осуществление действий, а не постижение истины.

Для представления схемы «деятельности» гомункулов предлагается вообразить тело, внешне неотличимое от человеческого. Голова, само собой, внешне так же неотличима. Хотя для функционализма это условие необязательно. Голова – как скала, в которой много-много са-

¹ Любую версию функционализма можно редуцировать к спецификациям столбцов, строк и отношений между ними: реляционная схема – универсальная модель любой предметной области.

мых разных пещер. Внутреннее строение этой «головы» забавно отличается от человеческой. Нейроны органов чувств соединены с системой освещения пустот в голове. В этих пустотах, как в пещерах, и живут гомункулы. В пещерные стены вмонтированы клеммы, подсоединяемые к моторно-выходным нейронам. Каждый гомункул должен выполнить простую задачу – реализовать «клетку» машинной таблицы, ячейки которой обозначают некоторое психическое состояние человека. На самой обозримой, центральной стене пещеры устроена доска объявлений. На доску прикрепляются открытки. В открытках написаны инструкции: какие действия должен предпринимать гомункул при получении данной открытки. Предположим, что внешнее относительно головы воздействие на чувственные нейроны приоткрывает тот или иной вход. Освещается какая-нибудь открытка, на которую попал луч света. На этой открытке (в адресной части) написано «G», т.е. она адресована гомункулам, которые занимают квадраты G (*G-мэнам*, как они сами себя могут назвать). Допустим, освещается открытка I, т.е. активизируются нейроны входа, обозначаемого посредством «I». Один из G-мэнов имеет возможность прочесть то, что написано на открытке I, так как она освещена. На открытке в содержательной части написано «M». Таким образом, когда в полость головы проникает луч света, соответствующий входу I, читается открытка с адресом «G», G-мэн нажимает выходную кнопку «M» и вызывает моторно-двигательные механизмы M.

Гомункульная машина Тьюринга может быть представлена конечной совокупностью четвёрок или пятёрок компонентов (пять составляющих требуется, если выход делится на две части – для изменения состояния системы и выдачи результата): 1) текущего состояния; 2) текущего входа; 3) следующего состояния; 4) следующего выхода. Каждый гомункул в определённый момент времени имеет задачу, соответствующую одной четвёрке (пятёрке). Как мы видим, особо изощрённого интеллекта и творческих усилий G-мэну не надо, да и активизироваться он может крайне редко – только тогда, когда луч света падает на адресованную ему открытку. Однако, несмотря на низкий уровень интеллекта каждого из гомункулов и при условии организации адекватной машинной таблицы, «сообщество» гомункулов способно реализовать самую сложную психическую деятельность, на которую только способны люди. Это происходит из-за того, что система гомункулов полностью моделирует логические состояния человека, т.е. функциональная

организация гомункулов эквивалентна функциональной организации человеческой психики.

Очевидно, что в «гомункульной голове» воспроизведены более ранние идеи Р. Кирка, в которых на роль гомункулов претендовали лилипуты. В работе, послужившей началом современных дискуссий по проблеме философских зомби – «Зомби против материалистов» [Kirk, 1974], – Р. Кирк описывает следующий эксперимент. Лилипуты атаковали Гулливера и захватили его голову. Получив доступ к нервной системе Гулливера, они отсоединили его нервы, стали отслеживать входную информацию, поступающую от афферентных нервов и посылать выходную информацию к эфферентным нервам. Поведение Гулливера при этом ничем не отличается от его поведения в исходном состоянии. Ему присущи все поведенческие диспозиции бывшего Гулливера, однако он ничего не ощущает, не чувствует, не владеет никаким феноменальным опытом. Гулливер стал зомби. Далее сценарий усложняется – голова Гулливера возвращается к исходному состоянию, лилипуты подключают нервы туда, куда следует, и покидают «поле боя». Возникает вопрос: «Станет ли Гулливер ощущающим существом?» По-видимому, нет, так как сомнительно, что чувства вернутся при восстановлении исходного состояния, ведь нам не известны причины, закономерно связывающие факты сознания с физическими фактами. В этом мысленном примере помимо всего прочего поднимается проблема определения пространственно-временных и материально-энергетических границ сопровождения психическими явлениями физических процессов. Но здесь – арена битвы философских зомби с искусственными личностями (см. подразд. 4.2).

Конечно, мысленные примеры очень похожи. Но очевидной новацией Н. Блока относительно «Гулливера» Р. Кирка – и это для нас крайне важно! – явилось привлечение идей машины Тьюринга. Мы существенно приблизились к машинному функционализму, в частности, и к компьютерной проблематике сознания, в общем.

Мысленный эксперимент «Китайская нация». В следующем мысленном эксперименте Н. Блока, получившем в исследованиях философии сознания собственное наименование «Китайская нация», изучается номологическая возможность функциональной организации сознания. Возникает претензия на своеобразное метафорическое «измерение» количества гомункулов (агентов, нейронов), востребованных для реализа-

ции психической деятельности человека. Идея такого рода квантификации поднимает вопрос о мощности множества элементов, из которых состоит машинная таблица, и «силы» устройств, предназначенных для её обработки. Допустим, полагает Н. Блок, достаточно миллиарда гомункулов, и предлагает рассмотреть население Китая, которое, как известно, достаточно многочисленно. Пусть каждый китаец снабжён мобильным телефоном для связи с каждым другим китайцем и искусственным спутником Земли (мы «воспользуемся» современными телекоммуникационными средствами, в отличие от Н. Блока, который применял рацию). Движения тел управляются радиосигналами, а сами сигналы посылаются в соответствии с инструкциями, которые китайский народ принимает со спутника. Сигналы со спутника доступны всем китайцам. Инструкции выдаются таким образом, что действующие по ним китайцы функционируют подобно отдельным нейронам. Средства радиосвязи играют роли синапсов. То есть в целом китайский народ дублирует нейродинамическую организацию человеческого мозга какого-то конкретного китайца.

Китайца мучают и избивают, снимая при этом томограмму мозга. Сигналы с нейронов больного мозга передаются через спутник соответствующим китайцам-нейронам. Каждый китаец, в соответствии с распределением болевых паттернов нейронов в мозгу индивида китаец-эталона, резко переходит в состояние боли, например, ранит себя кинжалом. Возникает вопрос в стиле социологического реализма – будет ли «китайская нация» в целом фактически испытывать боль, если в боли пребывают те китайцы, на которые указывают нейроны мозга индивида-эталона? Если имеется метафизическая возможность этого, то тогда психические качества (квалиа и феноменальные качества) оказываются соразмерными с функционалистской парадигмой.

Функционалистский ответ таков: да, китайская нация *может переживать квалиа*, например боль. Однако при этом каждый отдельный китаец не осознает того, что его нация испытывает боль, он «не видит за деревьями леса». Существо размером с нейрон (например, гомункул), находящееся в человеческой голове, может ошибаться в том, что психические феномены присущи только ему, а мозг в целом эти феномены не сопровождают. Система в целом, репрезентирующая *полный человеческий функциональный дубликат*, становится полным субъектом всех человеческих ощущений, или, по крайней мере, хотя бы дифференцированных ощущений, квалиа, как доказывает Шумейкер в [Shoemaker, 1975].

То есть система «китайской нации» фактически испытывает боль. Если такой подход верен, то тогда *ментальное имеет функциональную природу*, следовательно, феноменальные качества и квалиа специфицируются таблицей машины Тьюринга. Н. Блок – в оппозиции, но не в плане метафизических выводов, согласно которым индивидуальные сознания детерминируют либо не детерминируют общественное сознание. Метафизическое отношение к «китайской нации» возникает тогда, когда она рассматривается как система гомункулов, где каждый китаец – гомункул. Отсюда – неверные индуктивные заключения, которые В. Люкан назвал «гомункулизмом» [Лусан, 1987]. Н. Блок занимает оппозицию с формальной точки зрения, которая особенно ярко проявилась через два года в его концепции «*психофункционализма*» [Block, 1980] и сегодня считается классической функционалистской теорией сознания. Согласно этой концепции, функциональная интерпретация «психического» имеет континуальный характер. Поэтому репрезентативная мощность автоматной таблицы МТ слишком мала, посредством её не выразить структурного многообразия психических состояний.

Психофункционализм представляет закономерный этап развития функционализма МТ в плане акцента на более тонких моментах функционалистской трактовки сознания, нежели чем те, которые были предложены Х. Патнэмом в «грубой» машине Тьюринга. Подход Н. Блока сегодня имеет большое значение для проблематики сознания в общем и для развития идеи теста Тьюринга в частности, например, для создания «нового теста Тьюринга», реализуемого посредством так называемой «машины Блока». Психофункционализм и сопровождающие его блоковские новации заслуживают специального внимания. Они будут рассмотрены подробно в подразд. 2.2.

Следует отметить, что мысленные эксперименты Н. Блока не только развивают тематику функционализма МТ, но и имеют важные последствия для организации дискуссий по поводу социологического функционализма, соотношения репрезентативизма и коннекционизма, вычислимости ментальных феноменов на социальных системах и пр. Поднимаются традиционные философские проблемы реализма, концептуализма, номинализма. Главное то, что социокультурная тематика окрашивается идеями компьютеризма и, по сути, становится темой электронной культуры.

Тезис эквивалентности теоретического и машинного функционализмов

Попытаемся экстраполировать на проблематику сознания рассмотренный ранее тезис об эквивалентности различных способов формализации интуитивного понятия «алгоритм»: машины Тьюринга, частично-рекурсивных функций Чёрча и нормальных алгорифмов Маркова. Имеется большое разнообразие вариаций идеи о том, что сознание – это функция мозга-компьютера. Выбор конкретной разновидности функционализма представляется делом вкуса исследователя или выбором парадигмальной установки научной школы. Мои личные убеждения принадлежат инженерной интерпретации проблематики «сознание-функция компьютера» в силу признания того, что конструкторские формы научной деятельности в практическом отношении более значимы, чем теоретические, особенно когда теория аргументируется мысленными экспериментами метафизики.

В философии науки и техники многие являются сторонниками инженерного подхода. Например, полёты самолётов и космических кораблей стали возможны благодаря инженерному искусству, а не теоретизированиям по поводу полётов. Математик, возможно, унифицировал бы проблематику «мышления» посредством λ -исчисления – и здесь мы видим мощные инструментари апplikативной комбинаторики. Лингвист предпочёл бы «интеллектуальные» функции рассматривать как текстовые преобразования и т.д. Имеются интересные комбинированные методы, например, современный вариант теории алгоритмов, предложенный Грегори Чейтиным. Алгоритм раскрывается посредством языка LISP – языка обработки текстов, представленных в формате списков, прекрасно подходящего для λ -исчислений [Chaitin, 2003]. Короче, всё дело во вкусе исследователя. Наша скромная задача состоит в том, чтобы показать равнозначность различных подходов к функционалистской трактовке отношения «сознание/мозг (компьютер)», и, показав эквивалентность теоретического функционализма функционализму машины Тьюринга, предпочтение отдать второму как более демонстративному варианту, поскольку в практическом, инженерном плане он более перспективен. Однако в конце данного параграфа попробуем отказаться от машинного функционализма как от способа решения психофизической проблематики. Он очень далёк от проектов построения ре-

альных систем, которым можно приписать параметры Я, сознания, интеллекта. А раз так, то значимость всего современного функционализма подвергается сомнению, так как корни всех разновидностей функционализмов – в машинном функционализме. Поэтому делается вывод о необходимости реформирования функционализма. Для этого предлагается поставить субъекта функциональной системы в постнеклассическую позу и обозначить перспективы нового, тестового функционализма. Чёткая экспликация тестового функционализма будет предложена в следующей работе автора.

Раскроем тезис функционалистской эквивалентности путём демонстрации семантической тождественности языка, на котором описываются положения теоретического функционализма и языка описания работы машины Тьюринга. В этом нам поможет простая схема, предложенная Н. Блоком в небольшой работе «Функционализм» [Block, 1980a]. Работа, кстати, часто цитируется, причём иногда без ссылки на оригинал (например, [Вильянуэва, 2006, С. 138 – 140], хотя это представляется нам казусом переводчика). Индекс цитирования данной работы позволяет утверждать, что блоковская демонстрация функционалистской парадигмы стала общепотребимым объяснительным канон, раскрывающим особенности машинного функционализма.

Вначале Н. Блок обращает внимание на сравнения, которые постепенно позволяют сформировать представление о сознании как функции мозга. Например, «карбюратор» – функциональный термин. То, что им обозначается, смешивает воздух с топливом и подает полученную смесь в камеру сгорания. Так как карбюратор – инженерная конструкция, предназначенная для выполнения специальных функций, то термин «карбюратор» априори функциональный термин. Иное дело с другими терминами. Возьмем, например, термин «почка» (человеческая). В ряде случаев это – функциональный термин, так как обозначает то, что поддерживает требуемый химический баланс в организме, очищает кровь, выводя из неё вредные вещества. Имеются и другие способы употребления данного термина. Номинативный термин «почка», призванный называть обозначаемые предметы, ссылается на реальный объект в организме человека. Дескриптивный термин «почка» обозначает орган, находящийся там-то и там-то, причём на реальную почку он конкретно не указывает. Диспозиционный термин «почка» указывает,

например, на ментальное содержание переживания боли, в случае, скажем, острого нефрита.

Функциональная трактовка любой предметной области настолько универсальна, насколько универсален способ употребления функциональных терминов. А этот способ в самом деле обладает большой степенью общности. Так, согласно некоторым стандартным логическим системам, к функциональным терминам можно свести все другие термины. Например, номинативный термин предметной области – это функция с пустым аргументом. Дескриптивный термин получается путем подстановок функций в аргументы исходных функций и последующих функциональных преобразований. Диспозиционный термин может быть выражен предикатором, который обозначает функцию, определённую на отношении, и значение которой классическая логика соотносит с двумя объектами «истина» и «ложь». Неклассическая логика имеет большее количество значений предикаторов, включая, например, квантово-логические дескрипции, предполагающие состояние вероятностей «пробывания» значения в зависимости от инструмента означивания.

Таким образом, функционалистская трактовка любой предметной области представляется достаточно универсальной. Вследствие столь большой общности при употреблении слова «функционализм» следует весьма тщательно обозначать контекст его употребления. «Функция» в функционализме машины Тьюринга (МТ-представимая функция) отличается от функции, которая фиксируется бихевиористом в схеме «реакция-стимул». В бихевиоризме используется, как правило, диспозиционный термин, обозначающий намерение, желание, боль, уверенность и прочие пропозициональные отношения, которые «объективно» (точнее, контринтроспективно) оцениваются по входу и выходу системы. МТ-представимая функция отличается от функции в физикалистской схеме, когда утверждается отношение тождественности ментального и нейрального состояний – для отождествления требуется, по крайней мере, формальное отличие между двумя рядами феноменов. Отношение между этими рядами порождает функциональное представление. Для дуализма данное функциональное представление не требует последующего физикалистского редуктивного отождествления. Для идеализма редукция требуется, но, в отличие от физикализма, нейральное отождествляется с

ментальным. Чем же тогда функционализм МТ отличается от функциональных представлений нефункционалистских парадигм сознания? Тем, что МТ-представимая функция включает сведения о *состоянии* машины Тьюринга, в котором МТ оказалась в результате входного стимула и предыдущего состояния в контексте учёта всех других ментальных состояний и выходных реакций системы в целом. Согласно патнэмовскому функционализму, *логическое состояние МТ суть психическое состояние*.

Имеются вариации в трактовке физической реализации логического состояния. При разных аргументах (разных физических реализациях) имеем одно логическое значение (ментальное состояние), но не наоборот. В алгебре данное отображение принято называть «функциональным». Правда, в классической математической теории отображений накладывается ещё ряд ограничений по поводу сюръективности/инъективности/биективности и пр. Для нас эти усложнения не существенны. Остаётся заметить, что, возможно, апелляция к математической трактовке понятия «функции» собственно и послужила поводом назвать логику функционирования машины Тьюринга функционализмом МТ.

Вот, собственно, основная идея функционализма машины Тьюринга. В электронной культуре идея крайне притягательна. Появляется возможность отождествлять психику с алгоритмом, схемы нейральной динамики – с программой, а компьютер – с мозгом¹. Возникает соблазн решать посредством компьютерной технологии не только частные проблемы, связанные с материально-экономическим недостатком, политическим влиянием, социальным положением человека и пр. Ком-

¹ Формула «сознание/мозг» = «программа/компьютер» поддерживается Дж. Серлем, Н. Блоком. Однако она не совсем корректна. Программа – это специфически оформленный текст, реализуемый компьютером и выражающий идеальную, мыслимую конструкцию – алгоритм. Также и в психофизической проблематике мы видим различия между «сознанием» как: 1) феноменальным приватным качеством, 2) способом научного (и ненаучного объяснения) «сознания» и 3) механизмом мозговой активности. То есть здесь также просматриваются не два, а три члена отношения. Источник заблуждения – отсутствие компьютерной грамотности у философов, которые много пишут про «компьютерное сознание», но не пишут компьютерные программы.

пьютерная экспансия расширяется на мировоззренческие проблемы смысла жизни, смерти/бессмертия, общественного идеала и др.

Сегодня имеется много вариантов «функционализмов». Их количество равно количеству «функционалистов» и их оппонентов, помноженному на количество интервалов времени, когда исследователи придерживались определённой версии функционалистской идеи. Среди функционализмов немалой степенью общности обладает теоретический функционализм. Он, по сути, выражает народно-психологические концепции терминами функционализма МТ. Покажем равнозначность этих форм функционализма, предложив тезис эквивалентности:

Теоретический функционализм и машинный функционализм семантически эквивалентны

Проиллюстрируем тезис примером Н. Блока. Способ функционалистского описания ментальных феноменов специфицируется автоматной таблицей машины Тьюринга, распознающей чётное/нечётное количество единиц во входной последовательности:

Состояния Входы	S_1	S_2
1	«Нечётный» / S_2	«Чётный» / S_1
0	«Нечётный» / S_1	«Чётный» / S_2

У автомата два состояния – S_1 и S_2 , два входа – «1» и «0», и два выхода – слова «Чётный» или «Нечётный». Таблица описывает две функции. Первая от входа и выхода, вторая – от входа и состояния на следующем этапе работы автомата. Каждый квадрат кодирует два условных предложения, специфицирующих выход и следующее состояние при текущем состоянии и входе. Например, левый верхний квадрат кодирует конъюнкцию двух предложений: если машина находится в S_1 и видит «1», то она говорит «Нечётный» (т.е. увидела нечётное количество раз символ «1») и переходит в S_2 . Правый верхний квадрат показывает, что если автомат находится в S_2 и видит «1», то говорит

«Четный» и возвращается к S_1 ¹. Левый нижний квадрат показывает, что если автомат находится в S_1 и видит «0», то говорит «Нечетный» и остается в S_1 . Автомат должен начинать с состояния S_1 , так как, если его первый вход – «0» и он – в S_2 , то он неверно скажет «Четный». Данный недостаток Н. Блок исправил в следующей таблице:

Состояния Входы	S_1	S_2
1	«Нечетный» / S_2	«Четный» / S_1

Как и в предыдущем случае, автомат имеет два состояния (S_1 и S_2) и два выхода («Четный» и «Нечетный»). Различие в том, что вместо двух инпутов имеется всего один вход – «1», другие символы он не воспринимает. Как и прежде, в каждой ячейке таблицы имеем две функции, одна от входа и выхода, а другая – от входа и следующего состояния. Каждый квадрат кодирует два условия: состояние входа («1») и выхода («Четный»/«Нечетный»). Левый квадрат специфицирует, что если автомат находится в S_1 и видит «1», то говорит «Нечетный» (указание, что он видел нечетное число единиц) и переходит в S_2 . Правый квадрат показывает, что если машина находится в S_2 и видит «1», он говорит «Четный» и возвращается в S_1 . Этот автомат проще предыдущего, но выполняет ту же задачу, избегая ложного объявления нечетного числа. Теперь зададим вопрос: «Что такое S_1 »? Ответ: S_1 – это отношение, полностью задаваемое таблицей, т.е. определённое в контексте восьми отношений для первой машины и четырёх – для второй машины, по два отношения на каждую клетку. Дадим явную характеристику S_1 для второй машины:

Находиться в S_1 = быть в первом из двух состояний, которые связаны друг с другом, а также с входами и выходами следующим образом: пребывая в первом состоянии, при «1» на входе переходить во второе состояние и выдавать «Нечетный»; пребывая во втором состоянии, при «1» на входе переходить в первое состояние, выдавая «Четный».

¹ Термины, используемые для перевода с английского языка, звучат несколько странно для тех, кто знаком с теорией автоматов. Например, вместо «видеть» следовало бы употреблять «считывать», не «говорить 'Нечётный'», но «записать 'Нечётный'». Мы оставили оригинальную терминологию, которую использовал Н. Блок. Она подчёркивает, что понятие автомата применяется к ментальным состояниям, а не к техническим средствам обработки символов.

Далее Н. Блок применяет формат функционалистских суждений, инвариантных относительно предметной области. Формат был предложен Д. Льюисом [Lewis, 1965] (см. также в статье Дж. Левин [Levin, 1994]), однако, в отличие от наших представлений о возможности представления любого объекта и отношения предметной области функциональными терминами, здесь мы видим некоторое ограничение, по крайней мере, индивидуальные переменные и предикаторы функционально не выражаются:

Находиться в S_1 = быть x таким, что истинно $\exists P \exists Q$ [Если x находится в P и видит «1» на входе, то переходит в Q и произносит «Нечетный»; если x находится в Q и видит «1» на входе, он переходит в P и произносит «Четный»].

Остается сделать ещё один шаг – конкретизировать предметную область, т.е. перейти к проблематике сознания. Для этого предположим некоторую теорию ментальных процессов, которая характеризует все возможные соотношения между состояниями, входами и поведенческими реакциями. Пусть «боль» будет типовым ментальным процессом. Тогда можно сказать, что если вы сядете на стул, в котором торчит гвоздь, то это причинит боль и желание закричать «Ой». Функционалист следующим образом определяет «боль», применительно к представленному выше одновходовому автомату:

Пребывать в боли = быть в первом из двух состояний, которые связаны друг с другом, а также с входами и выходами следующим образом: сидя на стуле при наличии в нем гвоздя (вход), вскочить (перейти во второе состояние) и закричать «Ой».

Здесь Н. Блок конкретизирует широко известную методику Рамсея–Льюиса (см., например, в [Вильянуэва 2006, С.113 – 120]), специфицирующую ментальные состояния в терминах теоретического функционализма, т.е. посредством некоторого варианта языка символической логики, в данном случае логики предикатов первого порядка.

Пусть T – психологическая теория с n ментальными состояниями, из которых, допустим, 17-е является «болью». Тогда T можно задать следующим образом: $F_1, \dots, F_n$ – функциональные термы, которые характеризуют n ментальных событий; i_1, o_1 и т.д. – это характеристики входа/выхода:

Чувствовать боль = быть x таким, что:

$\exists F_1, \dots, \exists F_n [T(F_1, \dots, F_n, i_1, \dots, o_1, \dots)]$ и x находится в F_{17} .

Подобным образом функционализм *описывает ментальное нементальными (формальными) терминами*, которые характеризуют психические состояния без всякого упоминания о психическом содержании, обусловленном реализациями этих состояний.

Приведённые таблицы и их «ментальные» интерпретации демонстрируют следующее.

1. Психические состояния – то же самое, что и состояния машины (машины Тьюринга). Эти состояния задаются отношениями к другим состояниям, к входной стимуляции и выходной реакции. Например, боль – это такое состояние, которое *предопределяет* крикнуть «Ой» в контексте учёта всех возможных состояний.

2. Поскольку ментальные состояния – это состояния машины, то метод определения состояний автомата пригоден для объяснения и предсказания психических способностей человека. Ментальные состояния можно более-менее полно задать в терминах логикоматематического языка, описывающего психические состояния, чувственные данные и поведенческие реакции.

3. Функционализм устраняет главный недостаток бихевиоризма, так как, помимо внешне оцениваемого поведения, раскрывает внутреннюю структуры и механизмы, которые реализуют данное поведение.

4. Единичное функциональное состояние может быть представлено различными способами. Например, S_1 может быть механическим состоянием, состоянием электронной схемы, состоянием мозга и т.д.

5. Так как S_1 может быть реализовано различными способами, то неверно утверждение, что психическое состояние S_1 – это исключительно состояние нейронных клеток мозга: S_1 также является и электронным состоянием.

6. Так как первая и вторая таблицы взаимозаменяемы, то отсюда следует правомочность утверждения дизъюнкции логических состояний: идентичное ментальное свойство, качество, состояние может быть представлено разными таблицами.

Приведённые Н. Блоком положения позволяют явно сформулировать ключевое положение функционалистской концепции сознания (и мышления):

Существо может обладать сознанием (интеллектом), не имея мозга, т.е. психическое причинно связывается не только с состояниями мозга, но и с состояниями иного материального субстрата.

В приведённых блоковских утверждениях мы чётко видим редукцию теоретического функционализма к машинному функционализму. По аналогии с теоретико-алгоритмическими тезисами Чёрча – Тьюринга – Маркова, тождество различных разновидностей функционализма можно обозначить тезисом Н. Блока:

Спецификации теоретического и машинного функционализмов эквивалентны
--

Таким образом, парадигма функционализма в общем случае характеризует ментальное (и интеллектуальное, см. [Block 1981, Р. 8 – 10]) в формальных терминах функциональных отношений между значениями на входе, значениями на выходе и другими состояниями целостного организма. Сторонники машинного функционализма полагают, что реализация данной формальной системы вполне достаточна для имитации, моделирования, более того, фактического воспроизведения интеллектуального, ментального, психического. Главное – аппроксимативно полно осуществить формализацию психических феноменов и использовать мощные компьютерные средства. В принципе, под общий тезис машинного функционализма подпадает и информационный подход Д.И. Дубровского: сложная система кодовой зависимости между нейральными состояниями и феноменами сознания это суть функциональная составляющая самоорганизующейся системы. Чтобы постичь тайны сознания, самосознания, личности, Я с целью произвольной компьютеризации (манипуляции) состояниями собственного и «другого» сознания, следует расшифровать нейродинамические коды по аналогии с расшифровкой ДНК. Для самых сложных операций кодирования/декодирования информации в принципе достаточно формализма машины Тьюринга. Но хватит ли у машины Тьюринга мощности?

По всей видимости, поспешный марш-бросок через «объяснительную пропасть» «когниция/компьютер» не совсем верен. На перевале перед пропастью следует рассмотреть более конкретные варианты анализа когнитивных процессов. ФМТ даёт лишь слишком обобщённую систему спецификаций – спецификацию состояний. А как быть, скажем, с «событиями» или «содержанием» ментальных состояний? Необходимо более

чётко определяться с функциональной феноменологией сознания и, параллельно данной работе, искать способы построения инструментария, реализующего данные функции. Здесь возникает необходимость привлечения более поздней конструкции (относительно машины Тьюринга) – теста Тьюринга. Тест Тьюринга существенно дополняет функционализм (и теоретический и машинный) *феноменологией наблюдателя*, даёт концептуальную возможность построения системы функций, на основании которых специфицируется интеллект и, в более общем варианте, уже на базе привлечения нашей идеи (идеи комплексного теста) – специфицируются ментальные, но и биологические, персональные, социальные феномены.

В зарубежной аналитической философии сознания имеются многочисленные школы, направления, ответвления «функционализмов» компьютерной ориентации, столь важные для проблематики ИИ. В отечественной философии не так. По мнению автора, у нас функционализм не сложился ни в проблематике сознания, ни тем более в философии интеллекта. Показателем служит заседание специализированного круглого стола на конференции «Философия искусственного интеллекта» (Москва, 2005 г.) на тему «Функционализм и компьютерные науки» [Алексеев, Смирнова, 2005]. Мероприятием руководили Ю.И. Александров, Б.В. Бирюков, Д.И. Дубровский. В заседании Круглого стола приняли участие более 100 человек. С краткими вступительными докладами выступили Д.И. Дубровский и Ю.И. Александров. В этих докладах предпринята попытка очертить основные вопросы функционализма.

Д.И. Дубровский охарактеризовал современное толкование парадигмы функционализма, ее разные интерпретации представителями аналитической философии, в рамках которой интенсивно разрабатываются функционалистские подходы к решению проблем искусственного и естественного интеллекта. Он отметил, что большинство таких подходов носит редукционистский характер и связано с обоснованием концепции «сильного ИИ». Однако существуют нередукционистские варианты функционализма, которые совместимы с концепцией «слабого ИИ» и могут служить для продуктивной разработки проблемы сознания, теоретического объяснения природы субъективной реальности и создания перспективных направлений развития ИИ.

В выступлении Ю.И. Александрова были вынесены на обсуждение проблемы, касающиеся понимания соотношения психического и физиологического, влияния позиции так называемого «сильного ИИ» на стратегии конструирования ИИ. В частности, были сформулированы следующие вопросы: когда в эволюции материи возникло субъективное психическое отражение, кому допустимо приписывать наличие психики, как связана работа мозга и психика, как повлияет на деятельность создателей программ ИИ принятие положения о наличии у компьютера понимания, интенциональности, позиции первого лица, т.е. внутренних, субъективных ментальных состояний, и т.п.? Было описано системное решение психофизиологической проблемы, которое относится к двухаспектным теориям и может быть охарактеризовано как нейтральный монизм. Суть решения состоит в том, что психические процессы (характеризующие организм и поведенческий акт как целое) и нейрофизиологические процессы (протекающие на уровне отдельных элементов) сопоставимы не напрямую, а только через информационные системные процессы, т.е. через процессы организации элементарных механизмов в функциональную систему. Подчеркнуто, что нейтральный монизм и функционализм вполне совместимы. Предложено решение «трудной проблемы сознания» (как ее формулирует Д. Чалмерс): появление внутреннего мира связывается с тем, что рефлекторное, «запаздывающее» отражение, характеризующее неживую природу, сменяется опережающим отражением и целенаправленностью, характеризующими жизнь.

Дискуссия продолжилась на круглом столе «Философско-методологические проблемы когнитивных и компьютерных наук» в рамках IV Российского философского конгресса «Философия и будущее цивилизации», Москва, 24 – 28 мая 2005 г. (руководители – В.А. Лекторский, В.А. Глазунов, А.Ю. Алексеев). Обсуждения функционалистской проблематики ещё более чётко обозначили отечественные прорехи в знакомстве с зарубежными аналитико-философскими разработками функционализма.

Показателем синкретичности отечественного функционализма «сознания» явилось то, что *никто* из докладчиков не смог увязать функционалистскую трактовку сознания с компьютеризмом, не говоря уже о функ-

ционализме машины Тьюринга. За последние годы ситуация существенно не изменилась. Для отечественных исследователей функционализм до сих пор *terra incognita* и концентрированно не представлен ни в обзорных антологиях, учебных курсах, переводных изданиях, ни в тематике постоянно действующих семинаров, конференций. Тем не менее, на мой взгляд, отечественная наука способна внести оригинальный вклад в развитие функционализма. Однако это – отдельная тема исследований исторического плана, частично осуществленных в гл. 6.

Вернёмся к зарубежному функционализму, т.е. к тезису Блока. Заодно покажем перспективы тестового функционализма. Тезис вполне правомочен, так как: 1) функциональными терминами представимы все иные разновидности терминов – номинативные, дескриптивные, диспозиционные, предикативные; 2) программа редукции ментального к функциональному собственно и позволяет реализовать (по крайней мере, имитировать) функции интеллектуальной системы, не вдаваясь в подробности её онтологии.

Однако! В процессе проведения операции редукции, как правило, отбрасывается некоторый сопутствующий «осадок». При формализации психических качеств, состояний, свойств, отношений, процессов и пр. в «осадок» выпадает содержание, контент, квалиа. В результате редукции ментального к функциональному, аппроксимативно полного описания всех возможных входов / выходов / состояний системы из проблематики сознания изгоняется собственно «сознание», значение термина «сознание» переходит в область функциональной дескрипции. Вместе с этим перестают обладать качеством мои боли, надежды, мысли, моя свобода, вдохновение и, собственно, Я. «Сознание» утеряно. Его потерял теоретический ≡ машинный функционализм. Его потеряли все другие разновидности функционализмов, так как их оригинальная парадигма – это функционализм машины Тьюринга, в силу генетическо-концептуальных обстоятельств.

Наша задача – вернуть «сознание» человеку, изучающему «сознание». Здесь нет антиредукционистской программы организации отношений между ментальным и функциональным. Мы предлагаем кардинальное решение: покинуть стандартную схему рассуждений о психофизической проблеме и перейти в сферу постнеклассической методологии (В.С. Стёпин). Для этого следует разработать эвристически

более богатую теоретическую конструкцию, чем функционализм машины Тьюринга. Предлагается интегральная форма функционализма – *функционализм теста Тьюринга (тестовый функционализм)*. Полное изложение данной формы функционализма выходит за границы данной книги и требует более чёткого изложения в будущем. В настоящей работе мы лишь заложим основу тестового функционализма путем экспликации функций комплексного теста Тьюринга. Этому посвящены все последующие главы книги. Насколько полно нам является «возвращённое сознание» индивида компьютеризованного общества, определяется феноменологическими особенностями восприятия результатов функционирования компьютерной системы или, в общем случае, х-системы, как «интеллектуальной», «болевой», «волевой», «сознательной» и пр. Точнее, факт обладания «сознанием» задаётся убедительными, обоснованными атрибуциями данного факта. Система атрибуций строится наблюдателем – тьюринговым судьёй, т.е. Я.

Позиция тьюрингового судьи задаёт четыре типовых способа интерпретации компьютерно-реализованного «сознания», которые подразумеваются в тестовом функционализме. «Компьютерное сознание» – это: 1) *имитация ментального феномена* – лично я убеждён в том, что современный компьютеринг может претендовать лишь на этот способ; 2) *модель ментального* – к этому склоняется большинство представителей когнитивных наук, для которых некоторый механизм сознания – это действительная модель когниции, с присущими методу моделирования онтологическими притязаниями на её адекватность; 3) *репродукцирование ментального*, т.е. фактическое воспроизведение компьютерными средствами феномена сознания: уверенность исследователя базируется, например, на принципе инвариантности информации относительно её носителя и на том, что онтология «сознания» имеет информационный статус (позиция Д.И. Дубровского); 4) *креация ментального* – это не имеющие натуральных аналогов, новые феноменальные качества, свойства, события, продуцируемые компьютером, и которые не способны охватить человеческое понимание (сторонники сильного ИИ).

Мы очертили перспективы и кратко обозначили предмет тестового функционализма. А.М. Тьюринг претендовал исключительно на *имитацию* интеллектуального поведения. Наша задача намного амбициознее. С целью последующего решения этой задачи выдвинем тезис:

Реальные когнитивно-компьютерные системы, которым обоснованно атрибутируются Я, сознание, интеллект, возможны в контексте новой парадигмы функционализма – тестового функционализма (функционализма теста Тьюринга)

Тест Тьюринга – инженерная формализация интуитивного понятия «интеллект»

В отличие от машины 1936 г., предложенной для формализации интуитивного понятия «алгоритм», «машина Тьюринга» 1950 г. приобрела более конкретные очертания. Это – универсальная цифровая вычислительная машина (УЦВМ). Прототипом её послужила аналитическая машина Ч. Бэббиджа. Кстати, все современные широко распространённые компьютеры функционируют в соответствии с данным прототипом. УЦВМ лишь частично представима машиной Тьюринга, т.е. обеспечивает формализацию иного понятия, нежели чем понятие «алгоритм».

Для построения машины Тьюринга 1936 г. использовалась метафора работы математика при осуществлении вычислений. По всей видимости, А. Тьюринг описывал свою собственную работу математика-криптографа. В машине 1950 г. метафора другая. Идею этой машины А.М.Тьюринг разъяснял исходя из возможности выполнять все операции, которые могли быть выполнены *человеком-компьютером* (буквально «человеческим компьютером»¹). Человек-компьютер следует твердым правилам и не может от них отклоняться, однако сами правила меняются от одной ситуации к другой. Правила изложены в «Книге правил». Книга – это совокупность страниц, т.е. перфокарт. На каждой странице – своя собственная лента или совокупность лент (из предыдущей конструкции машины Тьюринга). Человек-компьютер, в зависимости от сценария своей деятельности, листает книгу и следует правилам, которые в ней изложены. Для человека-компьютера не ограничена поставка бумаги, на которой он осуществляет свои вычисления. В книге правил фиксируются сценарии, которые в общем случае не связаны с преобразованиями чисел. Для вычислительных операций предлагается вспомогательный калькулятор, хотя наличие его не существенно.

¹ В оригинале тьюринговой статьи используется выражение «human computer» – человеческий компьютер. Исходя из контекста суждений А.Тьюринга, «human computer» – это конкретное понятие. Поэтому перевод «человек-компьютер» предпочтительнее. Помимо прочего, выделяется компьютеринговый аспект антропологических представлений, наряду с другими аспектами, номинально закрепленными в историко-философской литературе: человек-животное, человек-растение, человек-машина и др.

Компьютер состоит из трёх частей: 1) *память* – хранилище информации в виде совокупности перфокарт неограниченной длины; 2) *исполнительный модуль* – это часть «человека-компьютера», выполняющая элементарные вычислительные операции; 3) *блок управления*, который принудительно выполняет команды (инструкции) из Книги правил. Таким образом, Книга правил – это усложненная «таблица команд» машины Тьюринга. Блок управления отслеживает, какая команда (инструкция) удовлетворяет условию запуска и установленному порядку выполнения команд. В нормальном режиме работы он выполняет команды в последовательном порядке. В других случаях встречаются инструкции безусловных и условных переходов. Также могут встречаться команды, задающие *вероятностный «скачок»* по таблице команд. В этом случае человек-компьютер, как полагает А. Тьюринг, проявляет «свободу воли», причём невозможно определить, наблюдая за операциями такого компьютера, имел ли место случай или же имел место закон, ведь эффект случайности может быть реализован генератором случайных чисел.

Прототип УЦВМ – аналитическая машина Бэббиджа – это полностью механическое устройство. УЦВМ – это электронное устройство, причём элементная база может быть весьма экзотической, скажем, память компьютера может быть построена на принципах сохранения акустических сигналов. Поэтому для человека-компьютера возможны не только дискретные, но и непрерывные способы функционирования¹.

В конструкции 1950 г. мы видим следующие новации: 1) возможна неопределённость действий (вероятностные скачки по таблице команд); 2) имеется возможность реализации уникального сценария, который после своего выполнения исключается из книги правил; 3) бесконечные ленты, которые бесконечно долго можно читать и столь же долго заполнять символами. То есть здесь мы имеем дело с другим понятием «алго-

¹ Современная элементная база способна создавать логические вентили, весьма различными и даже экзотичными способами. Помимо традиционных, собственно дискретных вентилях на полупроводниковой базе, имеются квантовые вентили (базовый элемент квантовых компьютеров), вентили на базе биологических молекул (биокомпьютеры), фононные вентили (фонон – квазичастица, представляющая квант колебательного движения атомов кристалла). Последние вентили сложно отнести к дискретным устройствам.

ритма», нежели то, которое поддерживалось машиной Тьюринга, так как проявляются следующие свойства: *индетерминированность*, *сингулярность* (может решаться лишь одна уникальная задача), *нерезультативность* (процесс может и не иметь остановки); *непрерывность*. С другой стороны, устройство человека-компьютера позволяет реализовать и стандартные свойства алгоритма – определённую, массовость (регулярность), результативность, дискретность.

Возникает понятие *квазиалгоритма* – «как бы» алгоритма. Термины, входящие в определение алгоритма (например, упомянутое выше определение по ГОСТ), получают модальную нагрузку и дизъюнктивную окраску. Например, квазиалгоритм – это неточное или недоопределённое предписание, либо описание, задающее или характеризующее вычислительный или исчислительный процесс, ведущий от варьируемых или уникальных исходных данных к искомому или неизвестному результату, либо от фактического результата к неизвестным исходным данным. Процесс, в принципе, может не завершиться. Возможны *клинчи* – такие состояния, когда процесс не прекращается в результате однозначности как минимум двух выборов. Собственно программа может быть набором *клюджей* – модулей, которые, как ни странно, работают, хотя программист не способен ни повторить процесс их создания, ни объяснить, почему они всё-таки работают.

С учётом выше изложенного примем краткую дефиницию:

Квазиалгоритм – это модально определённый алгоритм

Таким образом, тьюринговский цифровой компьютер 1950 г. – это формализация интуитивного понятия «квазиалгоритм». Для этого понятия не будем искать аналоги из математического и лингвистического подходов – инженерная трактовка наиболее демонстративна.

После того, как мы отследили математическую ветку, которая привела А. Тьюринга к идее «мыслящих машин», и, немного забегаая вперед, сформулируем компактный тезис:

Искусственный интеллект – это квазиалгоритмическая реализация теста Тьюринга

1.2. Игра в имитацию интеллектуального поведения: истоки дефинитной функции комплексного теста Тьюринга

Значение термина «тест Тьюринга»

Термин «Тест Тьюринга» появился в середине 1970-х годов – свидетельствует Дж. Лассаж в статье «Какую разновидность Теста Тьюринга подразумевал Тьюринг?» [Lassegue, 1988]. По всей видимости, впервые данный термин ввёл известный робототехник и редактор журнала «Машинное мышление» Д. Мичи в 1974 г. По крайней мере, Дж. Мур в статье 1976 г. «Анализ Теста Тьюринга» употребляет «тест Тьюринга» как вполне устоявшийся в научном сообществе термин [Moor, 1976].

Авторы статьи в известной Стэндфордской web-энциклопедии Г. Оппи и Д. Доу полагают, что термин «тест Тьюринга» («The Turing Test») в настоящее время используется в трех смыслах [Oppy, Dowe, 2008].

1. Чаще всего «Тест Тьюринга» используется для обозначения тезиса, высказанного А. Тьюрингом для решения вопроса «Может ли машина мыслить?». Как полагает Тьюринг, вопрос «может ли машина мыслить?» является совершенно бессмысленным в контексте рационального обсуждения. Однако вопрос проясняется, если его переиначить: «Может ли цифровой компьютер играть успешно в определённую игру – «игру в имитацию» («The Imitation Game»)?».

2. Выражение «тест Тьюринга» иногда используется для обозначения бихевиоральных тестов на наличие сознания, мышления, интеллекта в контексте исследований проблемы сознания.

3. Реже данное выражение используется в теоретических дискуссиях по определению логически достаточных условий на наличие сознания, мышления, интеллекта [Block, 1981].

На взгляд автора, имеются и иные значения термина «тест Тьюринга», непосредственно связанные с междисциплинарными исследованиями искусственного интеллекта. Исследователи философии искусственного интеллекта и компьютерно-ориентированных направлений философии сознания высоко ценят тест Тьюринга за возможность организации рациональных дискуссий по поводу идентификации

и реализации интеллектуальных систем различной природы. Он позволяет избежать вопросов, пустых и даже вредных с инженерной точки зрения, таких как «интеллектуальная сущность системы».

Дадим априорное рабочее определение «интеллекта» (в кантианском стиле), которое мы уже приводили вскольз в предисловии. Будем полагать, что примерно такого понятия придерживался А. Тьюринг, когда рассуждал об интеллектуальных машинах:

**Интеллект - это интегративная способность
упорядочивать чувственные данные, обоснованно рассуждать и
нормативно регулировать поведение**

Более конкретные теоретические суждения об интеллектуальности систем вызывают справедливые нарекания, приводят к неопределённости этого понятия и его непригодности для конструктивной аргументации. Независимо от метафизических достижений в обосновании «сущности» рассудка и разума современная инженерия интеллектуальных систем впечатляет. Техническая мысль согласуется с тьюринговой антиэссенциалистской стратегией идентификации интеллектуальности систем, различающихся по субстратным основаниям. Инвариантными относительно интеллектуальной способности, по Тьюрингу, являются люди и компьютеры¹.

Простая – потому гениальная – идея А. Тьюринга состоит в следующем: интуитивное понятие «интеллекта» можно конструктивно формализовать и реализовать посредством вычислительных машин в сценарии игры в имитацию интеллектуального поведения. Компетентный наблюдатель – судья этой игры – интуитивно понимает, что значит действовать «разумно», скажем, он пользуется при этом нашим априорным определением «интеллекта». Тестируя вербальное поведение неизвестной *x*-системы, т.е. задавая некоторые вопросы и получая на них ответы, судья оценивает её на предмет интеллектуальности. Неизвест-

¹ Сегодня физические инварианты «интеллекта» понимаются намного шире электронных УЦВМ. Это системы самой различной природы. Под тьюринговое тестирование попадают неорганические, биологические, социальные, глобальные, планетарные системы. Главное, чтобы их функциональная организация соответствовала сложности функциональной организации мыслящих существ.

но, является ли эта *x*-система человеком или компьютером. Судья принимает решение о наличии или отсутствии в ней каких-либо форм чувственной, рассудочной и разумной активности. Если *x*-система – это компьютер, то возникает инженерная задача её программирования. Предполагается возможность создания программы, способной ввести судью в заблуждение и вызвать у него иллюзию, что партнёром по диалогу является человек. Если результаты тестирования человеческой и компьютерной систем неотличимы, то утверждается, что компьютер способен играть в имитацию интеллектуального поведения, следовательно, способен мыслить.

То, что в контексте такой игровой ситуации компьютер «мыслит», не вызывает недоразумений. «Сущность» человеческого мышления не известна и, по-крайней мере сегодня, она непостижима. Аналогичная ситуация сложилась в авиации: универсальной теории полёта нет, однако, как ни странно, самолёты и космические корабли летают¹. Что такое «интеллект» – конкретно и вразумительно сказать невозможно, можно только воспользоваться общим априорным понятием, как мы предложили выше. Независимо от этой лингвистической проблемы искусственного интеллекта, сегодня компьютеры успешно распознают образы, выигрывают в интеллектуальных состязаниях, советуют при принятии решений, самостоятельно добывают полезные ископаемые в местах, несовместимых с человеческой жизнью и пр. Инженерная мысль и программистская практика значительно опережают теорию интеллекта, если таковая вообще возможна.

Изучим *антиэссеналистские истоки тьюринговой концепции*. Э. Ходгис, известный биограф А. Тьюринга, убеждён в том, что проблема интеллектуальных машин имела для автора теста паранаучную подоплёку – проблемой машинного интеллекта он стал заниматься в 1941 г. после знакомства с теологической книгой Д. Сейерс «Разум для

¹ Сравнение истории ИИ с развитием авиации имеет давнюю традицию. Недавно эту идею актуализировал Б. Уитби в выражении «аэродинамика интеллекта» в контексте социокультурных исследований ИИ. На русском языке издан перевод работы [Уитби, 2004]. К сожалению, перевод выполнен неквалифицированно (например, метод продукций – это, оказывается, «метод постановки»?!). Такая издательская работа вредна.

создания» (1941г.) [Sayers, 1941]¹. В книге проводится аналогия между человеком-творцом (в частности, сочинителем романов и пьес) и библейским учением о Троице. Любое человеческое творение состоит из Идеи (замысла человека-творца), Энергии (процесса записи идеи и реального «воплощения» на материальном носителе) и Силы (воздействия, которое произведение оказывает на аудиторию при её воспроизведении). Д. Сейерс полагает, что «троица человеческого творения» аналогична богословской Троице: Отец, Сын и Святой Дух.

В тьюринговой игре в имитацию нет таких конкретных аналогий, возможно, конечно, была некоторая «духовная» инспирация. Поэтому, по всей видимости, знаменитый биограф не прав. Так, в библиографии к своей статье Тьюринг ссылается, во-первых, на математические работы, в настоящее время ставшие классическими: А. Чёрча «Неразрешимые проблемы элементарной теории чисел» [Church, 1936], К. Гёделя [Godel, 1931], С. Клини «Общерекурсивные функции натуральных чисел» [Kleene, 1935] и на свою собственную [Turing, 1937], во-вторых, на технические работы: А. Лавлейс, в которой дано описание аналитической машины Беббиджа и Хартри «Вычислительные инструменты и машины» [Hartree, 1949], и в-третьих, на философские работы: С. Батлера (главы из его работы «Иерихон», названные «Книга о машинах» [Butler, 1865], Дж. Джефферсона «Разум механического человека» [Jefferson, 1949], Б. Рассела «История западной философии» [Russell, 1940]. Никаких паранормальных источников мы не прослеживаем.

Тем не менее, вопрос остаётся прежним: как можно создать такую машину, в которой *интеллектуальное возникает из совершенно неинтеллектуальных машинных операций*? В машинном функционализме, как мы видели, вопрос имеет расширенную трактовку: *как ментальное появляется из нементального (функционального)*?

В «до-тестовой» аргументации, т.е. до статьи 1950 г., такая «возможность» имела онтологические предпосылки [Turing, 1948]: 1) ин-

¹ Дороти Ли Сейерс (1893 – 1957) – широко известная в период между Первой и Второй мировой войнами английская писательница, автор детективов, поэтесса, драматург, эссеист, переводчик.

теллект продуцируется мозгом; 2) мозг представим работой конечного автомата; 3) любая функция мозга не только вычислима, но и реализуема конечной машиной, например машиной Тьюринга, у которой длина ленты ограничена, а то и вообще отсутствует (когда фиксируется лишь пребывание в некотором ментальном состоянии); 4) всевозможный спектр вычислимых функций, определённых на основе машины Тьюринга с бесконечной лентой, представляет лишь теоретический интерес, на практике следует учитывать финитность нервной системы – А.Тьюринг даже даёт приближённое значение ёмкости компьютерной памяти, необходимой для имитации работы мозга: 10^9 бит.

Но такой подход, идентифицирующий мышление с мозговой деятельностью, всё-таки не приемлем для инженера-математика: слишком много метафизически-тёмного. Наконец, в статье 1950 г. А. Тьюринг совершает *бихевиоральный поворот*. Несомненно, механическое лишено интеллекта. Возможности машины ограничены простыми и повторяющимися задачами. Однако правомочны два противоположных суждения: 1) машины предназначены для выполнения любого задания, которое может быть выполнено человеком, если он будет следовать своим инструкциям точно и абсолютно *бездумно*; 2) машины могут соревноваться с человеком в любой области интеллектуального труда. Если принимается первое, то тогда спор не имеет места. Человек работает как машина – и это стандартная схема работы человека на производстве или, скажем, при осуществлении преподавательской деятельности в современном бюрократизированном вузе. Здесь разум не нужен и вреден. Для второго суждения требуются конвенциональные оценки интеллекта по его проявлениям в лингвистическом поведении.

В итоге, посредством бихевиоральной игры в имитацию интеллекта, понятие «интеллект» полностью отделяется, во-первых, от квазирелигиозных убеждений о создании интеллектуальных машин как продолжения «творения Божья», и во-вторых, от метафизических притязаний компьютерного репродуцирования мозговой активности.

Лингвистический бихевиоризм выбран. Ответ на вопрос «Может ли машина мыслить?» теперь не заключается ни в выявлении «сущности» мышления, ни того, что подразумевается под словом «машина». А. Тьюринг предлагает поиграть в игру в имитацию интеллектуального поведения.

Этапы игры в имитацию

Игра в имитацию (**ИвИ**) происходит между мужчиной (**М**), женщиной (**Ж**) и судьей **С** (Interrogator)¹, ни пол, ни гендерная ориентация которого не имеют значения. Цель судьи – определить, кто из игроков женщина, а кто – мужчина. Задаются условия игры: 1) игроки изолированы от судьи непроницаемой перегородкой (стеной), чтобы по внешним признакам он не смог догадаться, с каким игроком имеет дело; 2) общение между судьей и игроками осуществляется посредством телетайпа (телеграфа); 3) реализуется вопрос-ответный режим: судья пишет вопрос, предаёт его по телетайпу, игроки письменно отвечают судье и передают ответы опять же по телетайпу; 4) вопросы могут быть на любую тему: от математики до поэзии, от погоды до шахмат; 5) женщина говорит правду и только правду, доказывая, что она – женщина, т.е. она помогает судье разобраться, кто из них кто; 6) мужчина – обманщик, его цель состоит в том, чтобы доказать судье, что он – женщина. Хотя мужчина, по желанию, время от времени может выражать истину.

На первом этапе игры в имитацию (**ИвИ1**) судья имеет дело с людьми, которых он не видит. Он должен определить пол игрока по серии вопросов, которые задает неизвестному собеседнику по ту сторону стены.

На втором этапе игры в действие вступает машина (компьютер) **К**. По задумке Тьюринга, новой постановкой вопроса «Может ли машина мыслить?» будет вопрос: «Что произойдет, когда машина будет играть вместо **М**². Будет ли **С** ошибаться столь же часто, как и в игре, где участниками являются только люди?» (см. 3-й абзац в [Turing, 1950, Р. 434]). Если судья будет ошибаться столь же часто, как и в **ИвИ1**, то Тьюринг считает, что **К** демонстрирует интеллектуальные способности.

¹ Interrogator можно перевести как «наблюдатель», «следователь», «опрашивающий». Нам мой взгляд наиболее подходящим словом является «судья», ведь у нас «игра», а того, кто оценивает игру, принято называть судьей. В дальнейшем чаще будет употребляться данный термин.

² У Тьюринга мужчина обозначается символом **A**, женщина – **B**, компьютер – **C**. Судья никак не обозначается. Так как у нас своя символика, на мой взгляд, совершеннее тьюринговой по причине способности формализовать многообразие частных тьюринговых тестов, то мы будем подставлять в цитаты свои обозначения, не оговаривая их при этом.

Данная ситуация изображена на рис. 1, который предложен в [Saygin, Cicekli, Akman, 2000]:

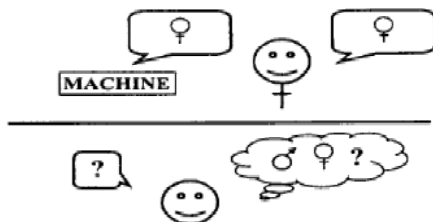


Рис. 1. Схема второго этапа игры в имитацию

Отметим следующий момент, который нам кажется важным для построения комплексного теста Тьюринга. Для изучения мысленных экспериментов способ графического изображения, приведённый на этой пиктограмме, не совсем удачен. Он нагляден, однако его трудно привести в соответствие с единой системой классификации и кодирования частных тестов. Использование этой системы позволит осуществить интегральное представление, программирование и применение частных тестов. Поэтому в следующей главе мы будем использовать формульно-символьный, а не графический способ записи теста Тьюринга и его различных модификаций.

Так вот, вроде бы со сценарием **ИВИ2** всё понятно. В самом деле, если **К** успешно играет роль **Ж**, то он, в самом деле, может мыслить, по крайней мере, мыслить по-женски. После достаточно подробного обсуждения устройства компьютера **К** как универсальной цифровой вычислительной машины А. Тьюринг задаёт вопрос по поводу «интеллекта» иначе, нежели чем в первом варианте: «...вопрос «Могут ли машины мыслить» должен быть заменён следующим: можно ли вообразить цифровой компьютер, который способен хорошо играть в игру в имитацию (т.е. хорошо подражать человеку – А.А.)? Можно задаться более общим вопросом: «Вообразимы ли дискретные цифровые машины, которые способны играть в эту игру?» (см. 2-й абз. в [Turing, 1950, P.442].

И далее сценарий неожиданно становится очень запутанным: «Обратим внимание только на компьютер **К**. Не кажется ли достовер-

ным то, что, увеличив его память и скорость до необходимых величин и снабдив его подходящей программой, компьютер **К** удовлетворительно сыграет роль игрока **М** в имитационной игре; *роль Ж при этом играет-ся мужчиной?»* [там же, курсив – А.А.].

То есть возникает новый сценарий – **ИВИЗ**. Женщина выбывает из игры. Вместо неё остаются «подражатели» – женщине подражает и компьютер и мужчина. Если компьютер, играя до этого роль **М**, по-прежнему обманывает, доказывая, что он – **Ж**, то что тогда делает **М**? А. Тьюринг не говорит ни о каких изменениях. Вспомним, что по начальным условиям (см. 2-й абз. в [Turing, 1950, P.434], роль мужчины заключается в том, чтобы доказать, что он женщина, но при этом он обманывает судью. Сейчас же, играя роль женщины, он должен доказывать, что он – женщина, но ведь он мужчина и должен обманывать. То есть, что тогда получается – мужчина поменял пол и стал женщиной? Или же он остался мужчиной, но тогда он может и обманывать и утверждать правду одновременно, т.е. противоречить самому себе, как известный лжец из парадокса Эвбулида. Короче говоря, совершенно непонятно, что имел в виду Тьюринг. Возникает какая-то абсурдная ситуация. Это достаточно тонкий момент, на который не часто обращают внимание комментаторы, переходя сразу к подражанию компьютером бесполого человека. Более тонкие наблюдатели – ими оказались в большинстве случаев социологи, занимающиеся гендерными проблемами, тщательно муссируют этот момент (по поводу Гендерного ТТ см. ниже).

«Зачем суетиться с женщиной, мужчиной и заменой?» – задаются вопросом Сэйджин, Цицекли и Акман [Saygin, Cicekli, Akman, 2000]. Ведь это значительно затрудняет понимание статьи. Неужели нельзя было сразу заменить пару «женщина-мужчина» на «человек-машина», как это сделали дальнейшие исследователи теста Тьюринга.

За этими недоопределённостями ИВИ, как они полагают, скрывается методологическая честность. Игра в имитацию как игра между полами, несмотря на то что она многими считается непонятной, – тщательно проработанная идея. Игра предоставляет объективную основу для сравнения: женщина выступает в роли нейтральной стороны. Поэтому оба обманщика – **М** и **К** – могут быть *оценены по степени правдивости в своей лжи*. Необходимостью *нейтральной* стороны и объяс-

няется то, почему в ИВИ три игрока, а не два – человек и компьютер. Именно благодаря третьей, нейтральной составляющей, судья может оценивать участников, играющих различные роли.

На мой взгляд, все эти трюизмы и противоречивые интерпретации по поводу трансформации полов не существенны. Можно полагать, что человек – это двуединство полов, условно разделённое эволюцией. У интеллекта с сексом каузальных связей мало. Можно так же оправдать запутанный сценарий предвидением одного из главных аргументов против мыслящих машин – «математического возражения» с позиции теоремы Гёделя, которая, как известно, является математическим выражением выше приведённого парадокса лжеца. Если **К** может играть роль «собственной роли», то значит, он преодолевает математическое возражение. Возможно, это так.

В любом случае разнообразие форматов ИВИ не мешает дискуссиям о тесте Тьюринга, в котором участвуют бесполые люди и компьютеры. Собственно, сам А. Тьюринг в последующих частях статьи «Вычислительные машины и интеллект» игнорирует транссексуальные проблемы и акцентирует на вопросе: «Может ли машина общаться на естественном человеческом языке?» (см. пример «беседы» в [Turing, 1950, Р. 434]: вопросы о поэзии, математике и шахматах вряд ли можно считать гендерными).

Игра в имитацию и канонический тест Тьюринга

А. Тьюринг заменяет вечный философский вопрос «Могут ли машины мыслить?» на «игру в имитацию интеллекта». Это – вариант лингвистической игры¹. Игра состоит из трёх сценариев (ИВИ1 – ИВИ3) и последующих модификаций (ТТ1 – ТТ4).

ИВИ1. Играют три человека – мужчина (**М**), женщина (**Ж**) и бесполой судья (**С**)². Игроки располагаются в разделённых непроницаемой

¹ Неизвестно, был ли А. Тьюринг знаком с концепцией языковых игр Л. Витгенштейна (1945 г.). В этих играх можно отметить много общих моментов, например, отказ от определения понятий, поиск значения слов в поведенческих сценариях, смысловой плюрализм.

² Даются обозначения, отличные от оригинала в целях лучшего восприятия. Также слово «Interrogator» было переведено «судьей» вместо принятого в [Тьюринг, 1960] громоздкого перевода «кто-нибудь задающий вопросы».

перегородкой комнатах, в одной – С, в другой – М и Ж. Диалог между ними ведётся средствами телетайпа. Цель С – определить посредством вопросов пол игроков. Вначале игры М и Ж – неизвестные Х и У. По условиям игры Ж говорит правду и только правду, помогая судьбе, а М может говорить и правду и ложь, стремясь обмануть С путем подражания Ж. Игра заканчивается, когда С утверждает одно из двух: «Х – это М, а У – это Ж» либо «Х – это Ж, а У – это М».

ИВИ2. Место Ж занимает компьютер К – универсальная цифровая вычислительная машина, на которой запущена программа типа анализа и синтеза связного текста. Возникает вопрос: «Будет ли С ошибаться столь же часто, как и в игре, где участниками являются только люди?». Этот вопрос и заменяет исходный – «Может ли машина мыслить?». Если судья будет ошибаться столь часто, как и в ИВИ1, то К демонстрирует интеллектуальные способности.

ИВИ3. Играют К и М. Мужчина М заменяет Ж. Цель С остаётся прежней – определить пол игрока. В такие игры, по мнению Тьюринга, можно будет играть в 2000 г. При наличии памяти 10^9 бит, достаточной скорости операций и соответствующей программы, у С будет не более 70 % правильной идентификации игроков, кто из них – М, а кто – Ж¹.

ТТ1: Играют человек Ч и компьютер К. Преследуются цели: для К – имитация Ч, для Ч – доказательство человечности, для С – решение вопроса: «Х – это Ч, а У – это К» либо «Х – это К, а У – это Ч».

ТТ2: Играет только К, цель которого обмануть С путем подражания человеку. Цель С состоит в различении искусственного и естественного интеллекта (кто с ним играет – Ч или К?).

Собственно варианты ТТ1 и ТТ2 в настоящее время принято называть *тестом Тьюринга* (ТТ). А. Тьюринг не предлагал этот термин. Это послужило поводом едкого утверждения о том, что А. Тьюринг стал наиболее известным в среде философов «компьютерной ориентации» за то, что никогда и не предлагал [Lassegue 1996]. В самом деле, при бук-

¹ Здесь следует возразить Дж. Лассажу [Lassegue, 1996], который указывает всего на два сценария игры в имитацию. Если следовать буквальной трактовке статьи, к чему призывает Дж. Лассаж, то можно достаточно утвердительно сказать, что Тьюринг описывает две фигуры в подразд. 1 своей работы, а третья фигура введена в конце подразд. 5 при описании универсальной машины.

вальном прочтении статьи, слово «тест» используется всего четыре раза, три из которых приходится на случай (особо интересный для нас), связанный с анализом возможности идентификации «другого сознания».¹ Однако, в защиту «теста Тьюринга» заметим, что, исходя из этических правил научной деятельности, не принято называть собственное творение своим именем. Вспомним, что А. Тьюринг не назвал собственным именем знаменитую «машину Тьюринга»: предложенная им дескрипция интуитивного понятия алгоритма, опубликованная в 1936 г. получила собственное наименование в журнале «Journal of Symbolic Logic» в 1937 г. [Lassegue, 1996].

Сформулируем ещё две формы теста Тьюринга, которые сегодня участвуют в междисциплинарных дискуссиях по искусственному интеллекту. Первая имеет некоторое отношение к тестированию интеллекта как главного атрибута *Homo sapiens*.

ТТ3: Судья тестирует некоторую систему **X** (**X** пребывает за тьюринговой стеной). Цель **C** состоит в определении человечности **X** либо в отсутствии у **X** такой характеристики.

Следующая трактовка выводит игру в имитацию за пределы дефиниции интеллектуального.

ТТ4: Судья тестирует некоторую систему **X** (**X** пребывает за тьюринговой стеной). Цель **C** состоит в определении наличия/отсутствия у системы когнитивных феноменов самого различного содержания: витальных феноменов (представляется ли система **X** живой или она мертва); ментальных феноменов (обладает ли система эмоциями, ощущениями, желаниями, убеждениями и пр.), персональных феноменов (можно ли оценивать **X** как личность, обладающую, по крайней мере, моральным вменением), социальных феноменов (**X** – это общество или бессвязный набор агентов)?

Несомненно, ТТ4 – наиболее общая трактовка игры в имитацию. Такая формулировка обычно фигурирует в современных когнитивных исследованиях.

¹ Дж. Лассаж ошибается, утверждая, что А. Тьюринг употребил слово «тест» два раза. На самом деле «тест» употребляется трижды в подразд. 6.4 при опровержении «Аргумента от “сознания”» и в подразд. 6.9 при опровержении «телепатического» аргумента.

1.3. Полемический стандарт теста Тьюринга: исток критической функции комплексного теста Тьюринга

Тьюринг во время написания статьи «Вычислительные машины и интеллект» [Turing, 1950], несомненно, чётко осознавал, что некоторые его идеи будут встречены враждебно, и, опережая, ответил на некоторые возражения, которым, как он полагал, его статья будет противоречить. Некоторые из них он обсуждал в более ранней статье 1948 г. (в [Turing, 1969]).

Отвечая на поставленный им же вопрос «Может ли машина мыслить?» Тьюринг неявно задаёт типовую структуру проведения дискуссий по данной теме, а также по другим темам, связанным с проблематикой искусственного интеллекта: 1) предмет возражения; 2) довод в пользу возражения; 3) ответ на возражение. В последующем исследователи ТТ, как правило, воспроизводят эту структуру и предлагают дискуссии (воображаемые и реальные) при описании частных тестов. Если же дискуссия явно не представляется, её можно реконструировать по многочисленным статьям оппонентов и пропонентов.

А. Тьюринг ограничился девятью положениями воображаемой дискуссии для защиты тезиса «Машина может мыслить!»). Эти положения предлагается назвать «*полемическим стандартом Тьюринга*». Предполагаемое возражение, опровергающее тезис Тьюринга, обозначено символом « $\rightarrow$ », довод в пользу – символом « $+$ ». Здесь нарушен порядок следования возражения, предложенный А. Тьюрингом в оригинальной статье. Возражения «теологическое», «антисциентистское» и «телепатическое» приведены в конце списка, как менее значимые для проблематики искусственного интеллекта. Некоторые положения *Я* сформулировал немного иначе, нежели чем в первоисточнике, но смысл от этого не поменялся. Каждый пункт тьюринговой дискуссии снабжён комментарием.

Математическое возражение

– Имеется несколько теорем, доказывающих ограниченность дискретных машин. Самая знаменитая из них – теорема Гёделя. Она показывает, что в замкнутой логической системе достаточной мощности обязательно найдется утверждение, которое нельзя ни доказать, ни опровергнуть, находясь в рамках этой системы.

+ Хотя и установлено наличие ограничений на мощность каждой конкретной машины, однако это утверждение берётся без всякого доказательства того, что человеческому интеллекту такие ограничения не присущи. Также возражения, построенные на таких теоремах, считаются не требующими доказательств, как и факт того, что машины, обладающие интеллектом, не ошибаются. Однако отсутствие или наличие ошибок не есть требование к мышлению [Turing, 1950, P. 445].

Математическое возражение апеллирует к гёделевской теореме и тьюринговым исследованиям описания формальных систем собственными средствами формализации. В контексте изучения интеллекта как активной человеческой способности данные исследования призывают «подняться» на уровень метаязыка и «над» алгоритмической рациональностью. Тогда будут разрешены логико-математические парадоксы, возникающие при аутоформализации, самоописании, самоидентификации и, в конечном счёте, вычислительной самореализации. Не только машине, но и человеку подобные парадоксы сложно разрешить. Например, парадоксы лжеца, брадобрея, теоретико-множественные парадоксы и пр. требуют от человека дополнительных ограничений – различать метаязык и объектный язык, фиксировать временные параметры относительно моментов утверждения о языковых и внеязыковых сущностях, вводить типы и дескрипторы в средства описания и пр. Такие достаточно жёсткие ограничения предлагает Б. Рассел, впервые за несколько тысячелетий достаточно близко подошедший к рациональному разрешению парадоксов. Предложенные Б. Расселом ограничения, позволяющие человеку не преодолевать парадоксы, а просто не впадать в них, ничуть не лучше условий традиционного программирования, в которых применяются строго типизированные типы данных.

Многие современные интеллектуальные и неинтеллектуальные системы программируются на базе нетипизированных языков, в которых отслеживание языковых уровней осуществляется за счёт учёта шагов рекурсивного определения термов, функторов и предикаторов предметной области. Всё это требует программной поддержки описания средств и способов программирования. Конечно, для преодоления парадоксов в общем случае представляется необходимым интуитивный выход за пределы формализованной рациональности. Но данная необходимость присуща не только компьютеру, но и человеку, затрагивает его внелогические, в частности, риторико-метафорические средства

«самостной» дефиниции. Хорошей иллюстрацией является пример с Мюнхгаузенем, вытаскивающим себя из болота за волосы (или за усы, в разных источниках по-разному).

В последующем данное возражение стали называть аргументом Гёделя – Лукаса – Пенроуза [Lucas, 1961; Lucas, 1996]. В нашей стране, несомненно, основным специалистом по данному вопросу является Е.М. Иванов – философ из Саратова. В его работах достаточно подробно разъясняется роль теоремы Гёделя в анализе возможностей машинного мышления [Иванов, 1999].

Прекрасный рецепт «снятия» математического парадокса предлагает В.Л. Васюков – надо перейти к такой логической системе, где теория Гёделя не работает. Очевидными преимуществами обладает паранепротиворечивая логика. В ней не имеет смысла доказательство путем сведения к абсурду, принятое в рамках классической логики, – данный метод был основой доказательства гёделевой теоремы о неполноте.

Возражение с позиции сознания

– Чтобы быть разумной, машина должна обладать сознанием (т.е. осознающей себя), чувствовать удовольствие от успеха, расстраиваться от неудач и т.д.

+ Следует использовать вежливую условность того, что все люди могут мыслить. Обычно по отношению к другим людям не принимается солипсизм. Также ТТ можно использовать для оценки факта запомненной информации, т.е. «действительно ли кто-то понимает что-то или заучил это как попугай. Загадки сознания следует решать не прежде, чем мы сможем ответить на вопросы о мышлении и, в частности, о мышлении машин» [Turing, 1950, P. 446 – 447].

Крайней точкой развития данного мнения выступает солипсизм. Единственный способ *действительно* узнать, мыслит машина или не мыслит – это *быть* машиной. Тем не менее, согласно данной точке зрения, единственный способ узнать мыслит или не мыслит другой человек – *быть* этим другим человеком. А это невозможно. Здесь возникает знаменитая проблема «другого сознания», значимость которой увеличивается по мере развития э-культуры [Алексеев, 2008-3]. Решение опирается на шаткий аргумент «по аналогии». Умозаключение «по аналогии» считается необоснованным и в большинстве случаев запрещённым логическим приёмом. Поэтому, чтобы не впасть в солипсические галлюцинации, А. Тьюринг предлагает вежливо согласиться с тем по-

ложением, что другие люди, как и «я», обладают способностью сознавать, и, следовательно, мыслить.

Таким образом, проблематика компьютерного сознания, по Тьюрингу, вырастает из проблематики ИИ. На мой взгляд, здесь возможны параллельные пути решения, в частности, на пути создания проекта искусственной личности, в рамках которой «интеллект» и «сознание» тесно переплетены. С другой стороны, убедительные концептуальные исследования (например, Фланагана и Полджера [Flanagan, Polger, 1995; 1998] аргументируют отсутствие корреляции между интеллектом и сознанием (см. далее Тест зомби).

Возражение о различных невозможностях

– Машины никогда не смогут сделать x , где x – любая человеческая способность – чувство юмора, способность влюбляться или обожание клубники.

+ Такая критика зачастую представляет собой замаскированный довод о сознании. Некоторые X неуместны в контексте ТТ как тесте на интеллект, например такие, как способность делать ошибки и любить клубнику [Turing, 1950, P. 448 – 450].

Как замечает Тьюринг [Turing, 1950, P. 449] такая критика обычно представляет собой замаскированный довод о сознании, т.е. вытекает из предыдущего возражения.

На самом деле здесь мы имеем не только x -невозможность, но и x -неизвестность. Сам человек до конца не раскрыл своих способностей. По всей видимости, если это окажется возможным, то это будет некий постчеловек. Конечно, компьютеры не способны реализовать некую x -возможность. Они не понимают юмор, не любят, не строят поэтических метафор и сравнений и пр. Но это и не нужно. Компьютеру должна быть отведена своя роль, «человеку – человеково, компьютеру – компьютерово». Не нужна машина, которая участвует в конкурсах красоты среди девушек. Очень чётко было подмечено данное возражение молодыми студентами МГУ им. М.В. Ломоносова в 1960-е годы, когда тема машинного интеллекта была новой и притягательной. Родилось стихотворение, которое приписывается Роману Солнцеву (записано со слов В.И. Самохваловой):

*Ненависти нету между нами
Умная, хорошая, прощай!
Не моргай зелёными глазами,*

Возражение леди Лавлейс

– Машины не способны к творчеству, не могут сделать ничего нового и не могут удивить нас. А если и удивляют своей способностью к мышлению, то в силу приписывания им этой способности со стороны человека.

+ Машины удивляли Тьюринга достаточно часто, как он это пишет. Признание чего бы то ни было удивительным требует достаточного *«созидательного мыслительного процесса»* независимо от того, кто – человек, машина или что-то еще – является автором этого удивительного и оригинального [Turing, 1950, P. 451].

В данной рубрике отрицается и утверждается креативная способность компьютерной системы. Возражение отводится, так как данная способность непонятна, невозможно найти подлинного автора идеи, не очевидны критерии оригинальности. К тому же результаты компьютерной обработки информации способны поразить любое воображение. Подробно аргументация рассмотрена при изучении теста Лавлейс.

Довод о неформальном поведении

– Интуитивно очевидно, что невозможно составить систему правил, которая описывала бы поведение индивида в каждой вообразимой ситуации. Если бы каждый человек имел набор правил на все случаи жизни, он был бы не лучше машины. Таких правил нет. Следовательно, люди не могут быть машинами.

+ Очевидно, что полный свод таких законов нельзя представить, всегда может быть упущено какое-то правило. Но, наблюдая за действиями машины, также невозможно предсказать её поведение [Turing, 1950, P. 452].

Этот аргумент тесно соприкасается с аргументом Лавлейс. Данный довод не проходит, так как поведение компьютера, решающего сложные задачи, предсказать невозможно, возникает слишком много алгоритмически неразрешимых вариаций. В принципе, имеются достаточно проработанные теории анализа программ, работающие посредством выделения из текста программы некоторых базовых алгоритмических инвариантов (схем программ). Однако они позволяют предсказать поведение программы при наличии чётко фиксированных значений фактических параметров. В интерактивном варианте, а именно такой предполагается в ИВИ, практически невозможны предсказания по поводу поведения программы: по текущим состояниям программы сложно

предсказать ответы на еще не заданные вопросы [Turing, 1950, P. 453]. И, наконец, по условиям игры, судья не имеет возможности осуществлять анализ программ, ведь он общается с х-системой.

Возражение по поводу непрерывности нервной системы

- Невозможно смоделировать поведение нервной системы с помощью дискретной машины, так как нервная система непрерывна.
- + Деятельность непрерывной машины можно представить в дискретной форме таким способом, что судья не заметит этого в условиях ИВИ.

Возражение опровергается возможностями аналогово-цифровых и цифро-аналоговых преобразований. Аналогично мультимедийной дигитализации аудио-видео-текстовых форм, в системах ИИ оцифровке подвергаются формы и способы интеллектуальной деятельности.

Возражение от боязни и неприязни («голова в песке»)

- В основе возражения лежит отвращение к идее о мыслящих машинах по причине того, что последствия появления и распространения таких машин будут ужасны. Большинство убеждено, что одной из главных особенностей человека является его способность к мышлению и эту особенность было бы неприятно разделять с машинами.
- + Возражение даже не стоит опровергать, а сторонникам следует найти какое-либо утешение, что-то вроде переселения душ [Turing, 1950, P. 444].

Большинство людей убеждено, что люди «особенные» среди других существ. Особенность заключена в способности к мышлению. Поэтому homo sapiens неприятно разделить эту способность с «железкой» – с машиной. Но научно-технологический прогресс не остановить! Знаменитой мифологемой является борьба компьютеров с человечеством, прекрасно продемонстрированная в фантастическом фильме «Терминатор». За это актер А. Шварценеггер был избран сенатором Калифорнии.

По поводу двух последующих тем полемического стандарта – теологического и телепатического возражения – имеется мнение, что они представляются лишними и не вписываются в научный дискурс. Так, авторы известной статьи в Стэнфордской энциклопедии вообще не стали их комментировать [Orru, Dowe, 2011]. Мне хочется восполнить этот пробел. Более того, я убежден в важности этих возражений – они выражают мировоззренческие интенции создания мыслящих машин. Первый – теологический аргумент – поднимает вопросы, связанные с творением и бессмертием. В рамках компьютерной тематики – это компьютерный креационизм и компьютерная танатология соответ-

ственно. Второй – телепатический аргумент – акцентирует на сверхчеловеческих способностях человека или, извините за парадоксальную формулировку, она инспирирована собственным антропологическим статусом человеческого существа.

Теологическое возражение

– Мышление есть свойство бессмертной души человека. Бог дал бессмертную душу только человеку. Следовательно, машины не могут мыслить.

+ В этом возражении сильно ограничивается всесильность всемогущего. Пытаясь построить мыслящие машины, мы поступаем по отношению к богу не более непочтительно, узурпируя его способность создавать души, чем мы это делаем, производя потомство.

Теологическое возражение отрицает способность человека создать мыслящую машину, так как только бог может одушевлять свои творения. Однако не надо ущемлять божественное всемогущество, создавшего человека по своему подобию, т.е. со способностью творить *искусственные* сознательные системы.

Проблема творения себе подобного от века вовлечена в дискурс техники. Каждая эпоха, обладая специфическими формами технического освоения социокультурной реальности, порождает инженерный миф творения. Бог слепил Адама из глины (Ветхий Завет). Галатея Пигмалиона («Метаморфозы» Овидия) была изваяна из мрамора. Гомункулус Вагнера («Доктор Фауст» Гёте) появился в результате химических опытов. Чудовище учёного – Франкенштейна («Франкенштейн» М.Шелли) – в результате опытов с электричеством. Шариков Преображенского («Собачье сердце» М.Булгакова) – в ходе нейрофизиологического эксперимента. Компьютерная инженерия породила робота Голема («Робот

и Голем» К. Чапека и «Голем» Г. Майринка)¹. Мифы о «творении» плюс некритично оцениваемые возможности компьютерной технологии создают, в лучшем случае, научные фикции, в худшем случае, паранормальный психоз. Для культурно-воспитательных целей такие мифы иногда полезны. Например, в невротическом поиске Пигмалионом целомудрия имплицитно представлены гендерные проблемы (потребность в искусственной женщине возникла из-за несоответствия реальности идеалу). Этическая проблематика просвещённого разума (Франкенштейн, Гомункулус, Шариков, Робот) убедительна для футурологических прогнозов. Однако, помимо культурологического интереса к памятникам литературного творчества, подобного рода компьютерные «мифологемы» онтологически и эпистемологически бессодержательны.

Проблема бессмертия в компьютерной окраске ничем не лучше и не хуже богословских исследований, например И. Брянчанинова² или, скажем, парамедицинских Р. Моуди³. Так, в компьютерной концепции Б.М. Полосухина [Полосухин, 1993] феномен «вечного бытия человека» задаётся самоприменимой машиной Тьюринга и содержит идеи самоорганизации и самовосстановления организма. Однако это – абстрактная идея, не имеющая оснований в реальности.

¹ См. Замаровский Войтех. Боги и герои античных сказаний: Словарь: Пер. с чеш. – М. Республика, 1994. – 399 с.; Гёте Иоганн Вольфганг. Фауст. Лирика: / Пер. с нем. / Вступ. статья и примеч. Ал.В. Михайлова. – М.: Художественная литература, 1986. – 767 с.; Шелли М. Франкенштейн, или современный Прометей. М.: Художественная литература, 1965 г. – 247 с.; Майринк Г. Голем: Роман / Пер. с нем. Д.Выгодского. – СПб.: Азбука, 2002. – 336 с.; Булгаков М.А. Собачье сердце. Повести, Рассказы, Фельетоны, Пьесы, Письма. – СПб., 1993 – 480 с. Впечатляющая подборка человеческих фантазий, повлиявших на научно-техническое «освоение» естественного интеллекта приведена Д.А.Поспеловым в [Поспелов 1982, С.5–42]

² Епископ Игнатий Брянчанинов. Слово о смерти. М.: «Р.С», 1991. – 317 с.

³ Моуди Р. Жизнь после жизни. М.: «Интерконтакт», 1990 – 95 с. См. также, например, «реинкарнационные» суждения про бессмертие в контексте «Мирового разума»: Мизун Ю.Г., Мизун Ю.В. Бог, душа, бессмертие. – М.: Экология и здоровье, 1992 – 336 с. Подобного рода литература встречается в большом количестве.

При изучении вечных мировоззренческих проблем – творения себе подобного и бессмертия – рациональная аргументация бессильна. Это – проблемы трансцендентного плана, выходящие за границы человеческого понимания.

В последующем теологическое возражение послужило концептуальным основанием для формулировки Креационистского теста Тьюринга как части Кибериадного теста (теста Швейцера). Рефлексия над этим возражением имеет важное методологическое значение в исследованиях «искусственных миров» [Макаров, 2005].

Телепатическое возражение («с точки зрения сверхчувственного восприятия»)

– Человеку иногда присуще сверхчувственное восприятие, а машине – нет. Человек способен интуитивно постичь правильный ответ на нечёткий вопрос судьи. Машина на это не способна.

+ Если считать, что телепатия возможна, тогда судью следует помещать в комнату, «защищенную от телепатии».

Некоторым людям, несомненно, присуще экстрасенсорное восприятие, интуиция, телепатия. Автор с этими интересными людьми встречался при реализации проектов по включению экстрасенсов в контур управления войсками в начале 1993 – 1996 гг. (заказчик – Генштаб Министерства обороны РФ [Алексеев, 2001]). Возражение «экстрасенсорности» снимается – компьютерные средства восприятия информации намного превышают ограниченные человеческие возможности в аудио- и видеодиапазонах, в иных формах – тактильных, гравитационных, осязательных, обонятельных. Для человека феномен экстрасенсорного восприятия представляется случайным и не очевидным. Если Вольф Мессинг заявлял о некотором сверхчеловеческом источнике своих экстрасенсорных возможностей, то не менее знаменитый экстрасенс и мастер психологических опытов Юрий Горный утверждает, что все его сверхчеловеческие способности – производное от длительных тренировок тела и интеллекта. Короче говоря, в силу неопределённости сверхчувственных способностей человека, «экстрасенсорные» способности компьютера также не определены. По поводу сравнения способностей Ю. Горного и компьютеров – см. ниже в контексте изучения «савантных» компьютерных систем.

Полнота и достаточность полемического стандарта А. Тьюринга

За полвека дискуссий по проблеме прохождения теста Тьюринга интеллектуальными компьютерами предложенная структура и содержание полемики превратились в своеобразный стандарт для исследователей. Каждый из них, обозначая свой собственный вариант теста, явно или неявно обращался к форме «возражение» / «ответ» (Н.Блок, Д.Деннетт, С.Уатт, С.Френч и др.). Вполне правомочен вопрос о полноте и достаточности полемического стандарта Тьюринга. Исчерпал ли А. Тьюринг девятью положениями всё многообразие возражений и ответов по теме «Может ли машина мыслить?». Полувековой ход дискуссий по ТТ показывает то, что дискуссии могут продолжаться настолько долго, насколько бесконечно разнообразие представлений о человеческом мышлении.

Из содержания полемического стандарта Тьюринга мы видим, что его тематика намного *шире* проблемы имитации интеллекта компьютерными системами. Первые семь рубрик непосредственно затрагивают проблему сознания, в двух последующих обсуждаются мировоззренческие проблемы «компьютерного» сознания. Поэтому, хотя оригинальный ТТ явно был определён как тест на интеллект, неявно, из контекста полемического стандарта сделаем следующий вывод:

Тест Тьюринга предназначен для изучения проблем компьютерной реализации когнитивных феноменов неопределённо широкого спектра

«Тест Тьюринга», получивший собственное наименование в конце 1960-х годов, сегодня трактуется по-разному и означает следующее:

- тезис о замене бессмысленного вопроса «Может ли машина мыслить?» на вопрос «Способен ли компьютер имитировать интеллектуальное поведение?»;
- логически достаточные условия диспозициональной или актуальной атрибуции (приписывания) когнитивных феноменов *x*-системе;
- лингво-бихевиоральный тест на выявление параметров когнитивных феноменов;
- критерий интеллектуальности *x*-системы;

- инструкция обмана человека в «человечности» компьютера;
- техническое задание на разработку интеллектуальной компьютерной системы;
- функциональная аппроксимация естественного интеллекта;
- спецификация естественно-языкового интерфейса с компьютером;
- проект компьютерного «усилителя» интеллекта;
- квазиалгоритмическая модель естественного интеллекта.

Имеется ряд других интерпретаций оригинального теста Тьюринга по поводу машинного «мышления», в которых всесторонне оцениваются инженерные проекты интеллектуальных компьютеров и критически обсуждаются мировоззренческие последствия их использования в обществе.

В аналитической философии искусственного интеллекта тест Тьюринга и его последующие модификации принято считать категориальными концепциями. Например, широко известная библиография Д. Чалмерса по философии ИИ открывается проблематикой теста Тьюринга [Chalmers Bourget, 2010]. Поэтому высокий уровень развития философии ИИ задаётся интегральным способом изучения и применения данного категориального аппарата.

Тьюринговая метафора интеллектуальной машины, привлекая всё большее внимание исследователей, служит концептуальным каркасом ИИ, является генератором новых идей, различных интерпретаций и толкований. Исследуются самые разные параметры теста: 1) длина – интервал времени, в течение которого осуществляется тестирование; 2) содержание тестовых вопросов; 3) «стена Тьюринга» – параметры скрытости игроков от судьи; 4) характеристики и особенности игроков и др. На многообразие новых трактовок влияет неопределённость идеи игры в имитацию, «чехарда» с ролями мужчины, женщины, машины и пр.

Сегодня насчитывается более тысячи крупных статей и специальных работ, в которых осмысливается ТТ. Поэтому стоит говорить о *тестах Тьюринга*, а не об одном единственном каноническом тесте, который, в свою очередь, не совсем чётко был определён самим Тьюрингом. Далее рассматриваются некоторые крупные модификации теста Тьюринга, которые мы назвали «частными тестами Тьюринга».

ГЛАВА 2. ЧАСТНЫЕ ТЕСТЫ ТЬЮРИНГА

Ряд исследователей тьюринговой проблематики (см., например, [Oppy, Dowe, 2008; Kirk, 2006]) полагают, что впервые на проблемы различения способностей человека и машины обратил внимание Рене Декарт в своей работе «Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках. Часть V» [Декарт, 1989, С. 250 – 296]. Он считал невозможным различие между нечеловеческим живым существом и машиной, если она ни внешним видом, ни внутренним строением не будет отличаться от животного. Так как человек обладает разумом, в отличие от других существ, то даже если машина будет иметь полнейшее сходство с нашим телом и поведением, различие возможно. Для этого следует протестировать машину: 1) на её способность пользоваться языком так же, как пользуемся мы, сообщая другим свои мысли – она провалит тест, так как нельзя представить, что она должным порядком будет располагать слова; 2) на её способность работать в самых различных контекстах – тест также не будет пройден, так как только человеческому разуму под силу охватить всё многообразие всевозможных поведенческих ситуаций, даже если в некоторых частных случаях машина может действовать лучше нас.

Назовем данные вопросы исследования поведения машины **«тестом Декарта»**.

Следует учитывать, что Декарт вводит понятие интеллекта (рассудка («*pense*»)) в терминах рефлексивного сознания или осведомлённости о себе (self-awareness). В «Первоначалах философии» (1640 г.) он пишет: «Под словом «мышление» («*рассудок*» («*pense*»)) я понимаю все то, что совершается в нас осознанно, поскольку мы это понимаем. Таким образом, не только понимать, хотеть, воображать, но также и чувствовать есть то же самое, что мыслить» [Декарт, 1989]. Т.е. тест Декарта – это тест на сознание, но не на интеллект.

Изменилось бы мнение Р. Декарта о возможности пройти такой тест машиной, если бы он познакомился с функционализмом машины Тьюринга и поработал с программами ИИ в среде Интернет? Следует, конечно, учесть уровень технических достижений 17-го века [Kirk, 2006]. Сегодня – в эпоху электронной культуры – А. Тьюринг положительно отвечает на вопрос о неотличимости человека и компьютера.

2.1. Становление частных тестов Тьюринга

В отечественной литературе отсутствуют переводные работы, аннотированные переводы и пр., посвящённые проблематике ТТ. Ряд крупных тестов (1980 – 2000 гг.) представлен в сборнике переводов [Алексеев, 2006-4]. Однако самые первые работы, в которых начинается обсуждаться проблема ТТ, для русскоязычного читателя остаются исторической редкостью. Тем не менее, первые дискуссии крайне важны, так как для современных исследователей ТТ может создаться впечатление, что мощные модификации ТТ, такие как тест Серля, тест Блока, тест Брингсйорда и пр. создавались на пустом месте. Но это не так. Остановимся подробнее на некоторых работах.

Интерпретация К. Гандерсона: тьюринговая игра в имитацию не соответствует «здравому смыслу»

Впервые на законность оценивания интеллекта в условиях проведения ИВИ обратил внимание Кейт Гандерсон (Keith Gunderson) в статье «Игра в имитацию» 1964 г. [Gunderson, 1964]. Выделяется два момента: 1) успешная игра в ИВИ может быть реализована без привлечения интеллекта; 2) интеллект – это нечто большее, нежели то, что может демонстрировать машина Тьюринга, т.е. тьюринговая ИВИ – лишь *единичный* пример из многообразия возможных компетенций разумного существа. Работа иллюстрируется развлекательными метафорами, которые собственно ориентируют дискуссии на «здравый смысл». Вначале задается вопрос «Могут ли камни подражать?» и описывается игра в «наступалки» («наступи на ногу») [Gunderson 1964, Р. 236], которая во многом аналогична тьюринговой игре в имитацию. Вновь играют мужчина (М), женщина (Ж) и судья (С). Судья С остается в изолированной комнате и не может ни видеть, ни слышать других участников. В стене – маленькое отверстие, через которое С может просунуть свою ногу. Игроки поочередно наступают на ногу С. Цель С – отличить мужчину от женщины, руководствуясь теми ощущениями, которые он получает, когда на его ногу кто-то наступает. Вместо цифровой машины Тьюринга предлагается коробка с камнями, висящая на стене над отверстием. Коробка снабжена видеодатчиками, направленными на отверстие в стене, и клапаном, который, открываясь, выпускает камни из коробки. При появлении некоторого предмета в отверстии стены, клапан открывается и камни падают. Судья просунул ногу в стену – камни посыпа-

лись. Теперь вопрос «Могут ли камни подражать поведению человека?» превращается в вопрос: «Что произойдет, когда коробка с камнями будет выполнять роль **М** в этой игре? Будет ли судья ошибаться столь же часто, как если бы игра велась между мужчиной и женщиной?» [Ibid, P. 236 – 237]. К. Гандерсон убеждён, что даже если коробка с камнями будет играть в «игру в наступалки» достаточно хорошо, то всё равно нет причин считать, что коробка подражает человеку. Единственный вывод, к которому мы можем прийти, это то, что коробка с камнями способна заменить человека исключительно в «игре в наступалки». Обобщение Гандерсона таково: «частью того, *что* происходит, является то, *как* это происходит (курсив мой – А.А.) [Ibid, P. 238]. В дальнейшем данный тезис усилится в аргументе Н. Блока против бихевиористской установки ТТ – описание диспозиции интеллектуального поведения, помимо спецификации возможных реакций на вербальную стимуляцию, должно включать механизм реализации диспозиции (см. далее «тест Блока»).

Второе утверждение Гандерсона – то, что о мыслительном процессе нельзя судить только по одному примеру. Успех компьютера в ИВИ это еще не достаточная причина для того, чтобы назвать его мыслящей машиной. Предлагается новое сравнение. Продавец пылесосов стремится сбыть свой продукт и всячески рекламирует пылесос, скажем, Шик-600, как многоцелевой. Далее он демонстрирует, как этот многоцелевой пылесос всасывает пыль. Покупатель спрашивает, что еще может делать пылесос. Удивленный продавец говорит, что пылесосы предназначены для всасывания пыли и Шик-600 делает именно это. Покупатель отвечает: «Я думал, он многоцелевой. Разве он не засасывает кусочки бумаги, травинки, грязь? Я полагал, что засасывание пыли – это лишь единичный пример того, на что на самом деле способен пылесос. Продавец отвечает: «Это пример того, что он делает. Он делает именно это – засасывает пыль» [Gunderson, 1964, P. 241]. Продавец, называя пылесос Шик-600 многоцелевым – ведь этот пылесос на самом деле многоцелевой, так как может пылесосить ковры, полы, стены и пр., – не смог продемонстрировать примеров других «компетенций» пылесоса. Точно так же, как термин «многоцелевой пылесос» подразумевает, что пылесос может что-то большее, нежели чем только засасывание пыли, у А. Тьюринга термин «интеллект» используется для обозначения более чем одной возможно-

сти, хотя имеется в виду лишь одна сторона интеллекта, востребованная для организации осмысленного диалога.

Ещё один пример – с бывшим прокладчиком туннелей, ныне безработным. Хозяин дорожно-строительной компании купил паровую дрель для того, чтобы сверлить дыры в скалах, поэтому рабочие-прокладчики оказались не нужны. То, что паровая дрель вытеснила рабочего – это факт, но данный факт не доказывает, что у машины есть мускулы, но лишь то, что туннели можно копать и без мускулов [Ibid, P. 254]. Соответственно, играть в тьюринговую ИВИ можно, не обладая интеллектом.

Возражение Дж.Г. Стивенсон – К. Гандерсон

Джон Г. Стивенсон в публикации «Об имитационной игре» [Stevenson, 1976] выступил против К. Гандерсона: машина, хорошо играющая в ИВИ, способна на многое. К. Гандерсон считал, что ответы, который даёт машина, даже если они согласованы с любым возможным текстом, в котором фиксируются вопросы, так или иначе демонстрирует узкий диапазон возможностей [Gunderson, 1964, P. 243]. В этом месте К. Гандерсон по сути воспроизводит тест Декарта. Дж. Стивенсон – против: даже если этот тест и не исчерпывает всего того, что связано с человеческим мышлением вообще, всё равно список «компетенций» компьютера очень внушителен.

Интерпретация Р. Пёртилла: тест Тьюринга как научная фикция

Ричард Пёртилл (Richard Purtil) в статье «Избиение игры в имитацию» (1971 г.) [Purthill, 1971] выступает с инженерной позиции, «главным образом не как философ, но как достаточно продвинутый программист» [Purthill, 1971, P.290]. Игра, несомненно, интересна, но это – некоторая разновидность научной фантастики. Нельзя вообразить то, что компьютер, играющий в ИВИ, будет разработан в обозримом будущем. ИВИ – это компьютерная мечта человечества. «Я съем электронную библиотеку у того, кто спроектирует компьютер, успешно играющий в игру в имитацию!» [Purtill, 1971, P. 293]. Если однажды компьютеры поведут себя так, как это описывается в научно-фантастической литературе, то тогда им можно будет приписать интеллектуальность. Но до тех пор всю компьютерную продукцию нужно считать результатом работы программиста. Даже если программа ведёт себя случайным образом (проявляя, по А. Тьюрингу, «свободу воли»),

всё равно невозможно вообразить методы и средства программирования, которые реализуют компьютер из ИВИ. Поведение разумных существ невозможно объяснить в терминах механики. ИВИ – это всего лишь поединок умов между интервьюером (судьей) и программистом: компьютер здесь – лишний, он совершенно не нужен.

Хотя первое в некоторой степени правдоподобно, то второй аргумент звучит неуверенно. Для его поддержания Пёртилл предлагает «чисто механические» аналогии: машины, сделанные из рычагов и колёс, которые могут выполнять те же задачи, что и компьютер. Если человеческий мозг – в высшей степени сложный комплекс таких механических процессов, то это означает, что «люди копируют не лучше компьютеров» [Ibid, P. 292]. Следует отметить, что данная аргументация восходит к мысленному эксперименту «Мельница Лейбница», в котором обосновывается онтологический статус духовных субстанций – монад. Если увеличить голову до размеров мельницы и войти в неё, то в её механических внутренностях никакого сознания мы не обнаружим. На наш взгляд, насколько непонятна аргументация Лейбница, который, отстаивая психофизический параллелизм, ведёт себя, как физикалист, измеряющий вес «души» в граммах, а её проявления – в монадах, настолько непонятно, почему машинные аналогии поддерживают аргумент Пёртилла. На момент написания статьи уже 10 лет уверенно существовал функционализм машины Тьюринга, предложенный Х. Патнэмом, основная идея которого состоит в отождествлении психических состояний от первого лица с логическими состояниями машины. Для функционализма в силу принципа множественной реализации совершенно не важен материальный субстрат, из которого сделаны игроки в ИВИ.

Возражение Д. Сампсон – Р. Пёртилл

Джеффри Сэмпсон (Geoffrey Sampson) в статье «В защиту Тьюринга» [Sampson, 1973] сомневается в аргументах Р. Пёртилла. Прежде всего, он полагает, что большинство ограничений, на которые указал Р. Пёртилл по поводу построения компьютеров, играющих в ИВИ, – это лишь временные преграды, которые могут быть преодолены, возможно, в не столь далёком будущем. Во-вторых, если поведение компьютера объяснимо в терминах его механизмов, то поведение человека объяснить не просто. Д. Сэмпсон также отводит утверждение

Р. Пёртилла по поводу того, что человеческая мысль – это настолько сложный комплекс процессов, что его невозможно сравнить с вычислительным процессом. Напротив, мышление проявляется в поведении, т.е. мыслительный процесс – это то, что человек делает. Поведение, в свою очередь, можно описать и соответствующим образом запрограммировать [Sampson, 1973, P. 593].

Интерпретация П. Миллара: тест Тьюринга – тест на компьютерное мышление

В небольшой (всего две страницы) статье «К вопросу об игре в имитацию» П. Миллар [Millar, 1973] смог рассмотреть несколько важных положений ТТ: 1) ИВИ не имеет ничего общего с тем, каким образом вовлечены компьютеры и люди в эту «запрограммированную» игру и вовлечены ли они вообще; 2) ИВИ не является правильным критерием оценки интеллектуальности машины, но лишь поводом для её «очеловечивания» – приписывания ей человеческих целей, морального облика, ментальности; 3) ИВИ – это не способ измерения уровня интеллекта машины, но лишь критерий оценки того, что машины обладают именно *человеческим* интеллектом. Поэтому следует быть непредвзятыми и полагать интеллектуальным каждое живое и неживое существо, пусть даже марсианина и машину, которое проявляет «поведение, хорошо адаптированное для достижения собственных и особенных целей» [Millar, 1973, P.597].

Согласно данной позиции, дефиниция компьютерного мышления не обязательно должна следовать представлениям о человеческом интеллекте и эволюции его интеллектуальных способностей.

Таким образом, в первые четверть века после появления идеи ТТ, и особенно в период 1960 – 1975 гг., проблематика ТТ состояла в критическом осмыслении игры в имитацию. Пионеры сразу обратили внимание на проблематичность включения эволюционно-эпистемологических идей в проблематику ИИИ, т.е. на внутреннюю способность системы к росту «знаний» и к усилению этой способности. Дискуссия пока ещё не вышла за пределы журнала «Mind», т.е. имеет место «междусобойчик».

2.2. Крупные версии теста Тьюринга

Тест Колби - параноидальный тест Тьюринга

Предложен К. Колби в работе «Искусственная паранойя» (1971 г.). Используется схема ТТ1. В роли **С** – врачи-психиатры, в роли **Ч** – параноики. Опыт тестирования показал, что почти в половине случаев психиатры не могли отличить ответы компьютера от ответов больных. На мой взгляд, успех **К** в прохождении данного теста обусловлен тем, что ответы невпопад – это правило психического отклонения, для реализации теста не требуется сложная техника, вполне приемлема лингвистическая недостаточность тестовых вопросов и ответов. По всей видимости, параноидальный тест могут успешно пройти программы генерации «постмодернистских стихотворений». Тест Колби будет подробно рассмотрен далее.

Социальный тест Тьюринга

Б. Шэннон [Shanon, 1989, Р. 31 – 32] впервые предлагает считать тест Тьюринга как тест для оценки социального интеллекта. В роли **С** выступает мать-природа (которую «не проведёшь!»), в роли игроков – сообщество людей, которые пройдут тест, если выживут. Тест интересен в концептуальном отношении – в плане идентификации и тестирования общественных форм сознания и «интеллектуальности» социальных систем.

Гендерный ТТ – тест на отличие мужчины от женщины

Впервые предложен Ю. Геновой в [Saygin, Cicekli, Akman, 2000, Р. 32 – 33]. Ю. Генова полагает, что ТТ служит решению вопроса «чем мужчина отличается от женщины», но никак не вопроса «может ли машина мыслить?». Имеются различные трактовки: 1) Тьюринг выражает склонность к транссексуальности¹; 2) Тьюринг – антифеминист [Genova, 1994], для него важно – «Может ли женщина мыслить?»; 3) Тьюринг, напротив, считает подражание женщине верхом совершенства; 4) Тьюринг просто шутил, так как в нём гармонично сочетались

¹ Тьюринг был обвинен в гомосексуализме, не выдержал тяжести обвинения и покончил самоубийством в 1954 г.

ум, юмор и секс [Hodges, 2002]. С. Яновская (1960 г.) [Тьюринг, 1960, С. 6] считает, что гендерный ТТ принципиально невозможен – судья будет правильно отличать женщину от мужчины в 50 %. В настоящее время гендерный тест является концептуальной основой общей темы «Секс и компьютер».

Тест леди Лавлейс – тест на творческие способности

С. Брингсйорд, П. Белло, Д. Феруччи в работе «Творчество, тест Тьюринга и (улучшенный) тест Лавлейс» [Bringsjord, 2001] продолжают возражение леди Лавлейс, предложенное А. Тьюрингом [Turing, 1950]. Тест задаётся следующим образом. Искусственный агент **К**, разработанный человеком **Р**, проходит тест Лавлейс (ТЛ) тогда и только тогда, когда: 1) **К** создаёт на выходе некий артефакт **А** (для леди Лавлейс – дочери Байрона – в качестве **А** выступала бы, по всей видимости, поэма); 2) **А** – это не результаты случайного стечения обстоятельств или сбоя машины. **К** всегда может воспроизвести (повторить) **А**; 3) **Р** не может объяснить, как **К** выдал **А**, даже при условии полной осведомлённости о **К** – структуре баз данных, алгоритмах функционирования и пр. Авторы считают, что: 1) ТЛ невозможно пройти путем обмана; 2) все существующие модели ИИ не пройдут ТЛ, так как разработчики знают их особенности, а «от себя» такие программы прибавить ничего не могут; 3) необходимым условием прохождения ТЛ является способность **К** самостоятельно модифицировать свою программу, т.е. **К** должен быть самообучаемым. Однако это условие не достаточно для прохождения ТЛ, так как правила самообучения не являются по сути оригинальными – они формируются в контексте «известного» для **Р**; 4) вариант наделения **К** «свободой выбора» путём моделирования, например мотивационно-волевых механизмов творчества, также не подходит, так как артефакт **А** будет представлен в некотором пространстве решений, что исключает его оригинальность.

Как известно, Тьюринг в ответ на возражение Лавлейс заменяет выражение «способность к творчеству» на «способность удивлять». Он считает, что любая ошибка в программном коде или аппаратный сбой «могут удивить». В ТЛ это исключается, так как: 1) условие «ошибки» оговорено в определении ТЛ и запрещено; 2) к **К** применяются жесткие условия контроля вычислительного процесса даже в случае сбоя. Тем

не менее, ТЛ не убедителен в силу неопределённости «творческой способности» и требований к артефакту.

Тест Лавлейс значительно востребован в тьюринговой проблематике и на протяжении данной работы он будет встречаться часто.

Тест Френча – тест на субкогнитивные способности

Тест предложен Р. Френчем в 1990 г. в работе «Субкогнитивные способности и границы теста Тьюринга» [French, 1990]. Р. Френч считает, что ТТ умело обошёл бескрайнее философское болото проблемы дух/тело, однако он не способен исследовать глубинные, подсознательные области человеческого интеллекта. Поэтому в ТТ следует включить так называемые субкогнитивные вопросы, т.е. любые вопросы, позволяющие идентифицировать низкоуровневую когнитивную структуру, например, подсознательную сеть ассоциаций (а эти вопросы, считает Р. Френч, неявно включает любой ТТ). Сеть ассоциаций состоит из нескётного количества статистически связанных друг с другом образов, которые воспринимались человеком в прошлом и которые могут быть активированы в любой момент. Какие именно ассоциативные ансамбли возникнут при некотором вопросе – зависит от социокультурной обусловленности и форм конкретной жизнедеятельности игрока; т.е. ТФ способен пройти лишь тот, кто живет и ощущает окружающий мир так же, как и люди. ТФ – это не тест на интеллект «вообще», а тест на социокультурно обусловленный человеческий интеллект.

Р. Френч предлагает ряд способов идентификации подсознательной сети ассоциаций, среди которых наиболее показательным, на наш взгляд, представляется *метод ассоциативных заучиваний*. Суть метода в следующем. У человека в ходе конкретной жизнедеятельности развиваются определенные ассоциативные связи между понятиями. Сила этих связей – величина варьируемая, и её можно оценивать временем, в течение которого происходит связывание одного понятия с другим (с понятием «хлеб» быстрее свяжется понятие «масло», нежели «собака»). Судья С может воспользоваться этим методом следующим образом: 1) за день до тестирования проводится контрольный опрос среди интервьюируемых (людей), подбирается произвольный набор слов и фиксируется статистически среднее время связывания понятий; 2) на следующий день при проведении теста судья задаёт кандидатам (среди которых **Ч** и **К**) то же самое задание, собирает результаты тестирования и

сравнивает с результатами контрольного опроса. Анализ позволяет достаточно чётко определить, кто из кандидатов – **К**, а кто – **Ч**. Человеком будет тот, чьи результаты окажутся более схожими с результатами контрольного опроса. Компьютер **К** будет постоянно проваливать ТФ, так как невозможно запрограммировать некий априорный способ ассоциативной связи понятий. Имеется единственный способ различения машиной всей данной совокупности ассоциативных связей между понятиями – это «окунуться» в поток конкретной жизнедеятельности людей, жить среди людей. Однако ТФ представляется неубедительным в силу ряда причин: 1) сомнительна возможность выявления значимых статистических зависимостей между понятиями на этапе контрольного опроса; 2) разработчик **Р** может пользоваться различного рода словарями, где учитывается лексическая сочетаемость слов, например [Мельчук, 1995].

Тест Блока – психофункционалистский тест

Н. Блок в работе «Психологизм и бихевиоризм» [Block, 1981] критикует канонический ТТ, показывая, что «игра в имитацию» является разновидностью оценки поведения в контексте лингвистического бихевиоризма. Так как предметом бихевиористского анализа является не ментальная структура, а объективно фиксируемые параметры поведения (реакции), определяемые внешними воздействиями (стимулами) (в лингвистическом бихевиоризме стимулы и реакции – это вербальные выражения), то здесь имеется ряд стандартных антибихевиористских возражений на возможность осуществления адекватной оценки: 1) *аргумент Чизголма и Гича*, согласно которому невозможно выделить конкретное психическое состояние в поведенческой диспозиции без учёта *всей совокупности иных психических состояний*. Анализ каждой реакции может продолжаться бесконечно в континууме рассмотрения внутренних процессов, и результаты данного анализа невозможно зафиксировать в схеме «стимул-реакция»; 2) *аргумент «совершенный актер» Х. Патнэма*, согласно которому *различные классы психических явлений могут продуцировать одинаковые поведенческие диспозиции* и, наоборот. Например, в силу принятых в обществе законов «сверх-супер-спартанцы» не будут выказывать боль, даже если на самом деле люди её испытывают. Напротив, человек, не испытывающий боли, может притворяться, что ему больно; 3) *аргумент «паралитики» и «мозги*

в бочке»: эти «пациенты» вообще не выказывают определенных признаков, означающих, что они испытывают боль.

В свою очередь, интеллектуальное поведение, как правило, продуцируется комбинацией: {интеллект + обычная склонность к мышлению}. Однако невозможна комбинация {отсутствие интеллекта + желание обмануть, что интеллект имеется}.

Отсюда вытекает: 1) нужно изменить ТТ, так как он рушится под натиском антибихевиористских аргументов; 2) следует искать иное определение ТТ.

Предлагается новая формулировка ТТ (нео-ТТ), или тест Блока, который означает *способность продуцирования* осмысленной последовательности вербальных реакций на некоторую последовательность вербальных стимулов, но *не акт продуцирования*. Новая постановка усиливает ТТ, ТБ пройти намного сложнее, так как снимаются вышеперечисленные аргументы: 1) можно представить много факторов того, что интеллектуальная система *не будет расположена* к выдаче *осмысленных* ответов, однако эти факторы не скажутся на её *способностях*; 2) невозможно притворяться разумным и не быть таковым; 3) у «паралитиков» и «мозгов в бочке» имеется *способность* реагировать осмысленно, однако им недостаточно средств для реализации.

Для прохождения теста Н. Блок предлагает машину, которая способна производить осмысленную последовательность вербальных реакций на вербальные стимулы [Block, 1981, Р. 9 – 13]. Поведение машины полностью предопределено людьми – проектировщиками, инженерами, программистами и др. Разработчики упорно трудятся, применяют свой разум для реализации всевозможных последовательностей осмысленных ответов на предполагаемые последовательности вербальных стимулов. На поведенческом уровне данная машина выглядит интеллектуальной. Однако знание её внутреннего устройства убеждает в полном отсутствии у неё «интеллекта» – *интеллект, «проявляемый» машиной как в форме актуального поведения, так и в форме диспозиции интеллектуального поведения – суть интеллект разработчиков.*

Н. Блок, по сути, подсказал возможность прохождения ТБ – надо включить в сценарий имитационной игры разработчика Р. Тогда на любой вопрос со стороны С может быть получен ответ со стороны К при содействии Р. Единственный способ для С различить К от Ч (и тем са-

мым дискредитировать компьютерную программу) можно назвать «анатомическим» – т.е. использовать принцип «вскрытие покажет». Только тогда программа не пройдёт тест Блока, когда вскроются внутренности устройства реализации.

Тест Серля - тест на «понимание»

В 1980 г. Дж. Серль предлагает знаменитый мысленный эксперимент «Китайская комната» в работе «Разумы, мозги, программы» [Searle, 1980]. Здесь отрицается концепция сильного ИИ, согласно которой при прохождении ТТ компьютер не просто моделирует человеческую способность к пониманию (это характерно для позиции слабого ИИ), но и: 1) на самом деле *понимает* текст; 2) компьютер и её программа объясняют человеческую способность «понимания».

Сценарий Дж. Серля разыгрывается по схеме ТТ1, но теперь роль **Ч** и **К** совмещает в одном лице сам Серль-в-комнате (Searle-in-the-room) – мысленный эксперимент осуществляется с позиции первого лица. Судей двое – китаец и англичанин. Когда задаёт вопросы судья-китаец, то Серль-в-комнате, играя роль **К** и выполняя чисто синтаксические преобразования над китайскими иероглифами (он снабжен инструкциями перевода на родном английском языке), выдаёт китайцу ответы, которые тот воспринимает как осмысленные. Когда вопросы задаёт судья-англичанин, Серль-в-комнате играет роль **Ч**, непосредственно понимая задаваемые вопросы и не выполняя никаких синтаксических преобразований. Отличие между первым контуром игры (когда понимания нет) и вторым контуром игры (когда понимание очевидно для Серля-в-комнате) заключено, по мнению Серля, в *интенциональности*, так как человек – это биологическое существо, способное ощущать, действовать, понимать, обучаться и т.д. Способность к интенциональности заложена в биологическом прошлом человека. Интенциональные же феномены (в частности, квалиа) не редуцируемы к формальной схеме. Формальная модель по причине отсутствия самостоятельных казуальных сил не самодостаточна для воспроизводства феномена интенциональности. Интенциональность, якобы «имеющаяся» у **К**, на самом деле имеется лишь в головах **Р** (кто их программирует) и **С** (кто интерпретирует ответы).

Детально тест Серля будет раскрыт в составе одного из вариантов комплексного теста, поэтому здесь не следует на нем останавливаться.

Сделаем лишь ряд замечаний. Несомненно, данный тест имеет важное значение для развития когнитивной науки. Ряд исследователей полагает, что именно ТС послужил становлению нового этапа развития когнитивной науки – в целях опровержения ТТ [Cole, 2009]. Хотя условия прохождения ТС по сути воспроизводят условия прохождения ТБ, так как здесь также концептуально задействован **Р**. Поэтому тест Сёрля пройти можно, ничего принципиально нового относительно ТБ здесь нет.

Тест Ватта – инвертированный ТТ

Стюарт Ватт в работе «Наивная психология и инвертированный тест Тьюринга» [Watt, 1996] считает, что успех прохождения ТТ связан с естественной, наивно-психологической установкой приписывать ментальные свойства любым системам – другим людям, животным, природным явлениям, в том числе и компьютерам. Для преодоления наивной позиции наблюдателя предлагается инвертированная версия теста Тьюринга. Это означает, что надо «обратить» роли игроков. *Судья должен быть судим*, так как именно он приписывает интеллект системе, которая таковым может и не обладать. То есть теперь тестированию подлежит судья. Для этого схема ТТ2 модифицируется – вместо одного судьи теперь двое – судья-компьютер (**С-К**) и судья-человек (**С-Ч**). Тестирование осуществляется как прежде. Но теперь **С-Ч** (в роли его должен выступать опытный эксперт) должен сличать ответы от игрока **К** или **Ч** с показаниями **С-К**. Если **С-К** начинает приписывать интеллект тестируемой системе, то это является поводом для индуктивного заключения о том, что **С-К** – машина.

По поводу невозможности организации «инвертированного» тестирования существует ряд убедительных мнений: 1) тесту присуща *проблема дурной бесконечности*, так как для каждого **С-Ч**, судящего **С-К**, должен стоять другой **С-Ч'**, который судит первого **С-Ч**, и т.д. [French 1995]; 2) тесту присуща *проблема параллелизма*, так как невозможно обнаружить момент, когда именно **С-Ч** начинает отличать себя от **С-К** [Bringsjord, 1996]. Таким образом, вопрос о возможности прохождения ТВ должен решаться после вопроса о самой возможности тестирования.

Тест Харнада – тест НБИКС-системы

Многоуровневую классификацию тестов Тьюринга предлагает С. Харнад в работе «Разумы, машины и Тьюринг: неразличимость неразличимостей» (2001 г.) [Harnad, 2001]. Выделяются следующие уровни.

ТХ0¹. Это – уровень «игрушечных» ТТ – не полноправных тестов, а лишь некоторых фрагментов, ограниченных как по длине, так и по содержанию. Такие тесты не отвечают исходному замыслу А. Тьюринга. Однако все попытки моделирования интеллекта, известные на сегодняшний день, выше данного уровня не поднялись.

ТХ2 – общепринятое понимание ТТ. Именно его и имел в виду А. Тьюринг. Иногда ТТ данного уровня называют тестом «друг по переписке». Длина **ТХ2** равна протяжённости человеческой жизни С. Программа «Элиза» из примера «Девушка по переписке» принадлежит этому классу программ.

ТХ3 – это так называемая «роботизированная» версия ТХ2. Имеется возможность манипуляции предметами внешнего мира. Процедура идентификации систем, которые проходят данный тест, требует реализации принципа «Вскрытие покажет».

ТХ4 – это компьютерные системы, неотличимые как в плане ТХ3-неотличимости, так и в плане микрофизической организации системы. Здесь имеет место «тотальная неотличимость» компьютерной системы от человека, включая мельчайшие внутренние нюансы телесного строения. На данном уровне можно разместить так называемый «Тест Тьюринга с мозгом» («Brainy-Turing-Test»), где в дополнение к поведенческой неотличимости требуется неотличимость на уровне нервной системы и предполагается использование коннекционистских моделей [Flanagan, Polger, 1998].

ТХ5 – на этом уровне происходит отбор ТХ4-систем, соответствующих Единой Теории Реальности, которая способна (если она вообще возможна) объединить все возможные знания о природе, обществе, человеке. Данный уровень С. Харнад ввёл, по всей видимости, для того, что-

¹ Мы вводим свои символы для обозначения уровней, отличных от харнадовских, с целью достижения единообразия нашей классификации ТТ.

бы допустить возможность истинности нескольких отличающихся друг от друга теоретических описаний одной и той же реальности (т.е. ТХ4-систем). Истинность в данном случае имеет различные метафизические предпосылки у создателей различных теорий.

В соответствии с классификацией С. Харнада, современный уровень достижений ИИ не превысил «игрового» уровня ТХ0. Однако то, что прохождение ТТ возможно, пускай даже на этом уровне, демонстрируют факты. Рассмотрим эти факты.

2.3. Возможности прохождения тестов Тьюринга

Практическая возможность (тест Лойбнера)

В январе 2000 г. в Dartmouth College состоялась юбилейная конференция «Будущее теста Тьюринга: следующие пятьдесят лет», на которой обсуждались такие теоретическими вопросы, как оценка влияния ТТ на современность и определение прогнозов программирования ТТ на будущее. Практическая часть конференции состояла в розыгрыше сценария ТТ1. Участвовали шесть программ. В целом результаты не подтвердили предсказаний Тьюринга – судьи правильно отличали компьютер от человека в 91% после 5 минут и в 93% после 15 минут соревнований [Platt, 1995]. Тем не менее, в настоящий момент в сфере разработки ТТ-программ сформировалась своеобразная индустрия, в основном игрового характера (уровень ТХ0). С 1991 г. введена специальная премия, соразмерная Нобелевской – Лойбнеровская премия, которую учредил Хью Лойбнер (Hugh Loebner). Премия в 100 000 долларов присуждается за выигрыш, 2 000 долларов – за участие. Установлена длина теста – три часа. Тестирование проводится по сценарию ТТ1. На вопросы десяти судей отвечают пять **К** и пять **Ч**. Пока ещё нет победителя, которому можно было бы присудить полную премию [Алексеев, 2006-2].

Однако не следует отождествлять «игру в имитацию», которая прослеживается на состязаниях Лойбнера, с тестом Тьюринга. Тест Тьюринга – это теоретический план исследований искусственного интеллекта, не связанный с практической реализацией. Прошла или не прошла программа тест Лойбнера – это аналогично тому – пришла ли данная лошадь на скачках первой или вообще не пришла к финишу. То

есть тест Лойбнера не гарантирует даже эмпирической эмпирической валидности теста Тьюринга.

Совершенно иначе следует оценивать серьёзные программы искусственного интеллекта. Например, под тьюринговое тестирование подпадает любая экспертная система, система распознавания образов, генерации логических теорем и пр. Все они – кандидаты на прохождение, и, более того, все эти системы с учётом достаточных ограничений предметной области могут проходить тест Тьюринга.

Попробуем более скрупулёзно разобраться с тем, что значит «возможность» прохождения ТТ.

В контексте проблемы сознания, философии искусственного интеллекта, в любой теоретической сфере, в которой используются мысленные эксперименты, к слову «возможность» следует относиться осторожно. Здесь, конечно, не приемлемы жаркие диалектические дебаты на тему «возможности – вероятности – действительности», с характерными для них тёмными высказываниями типа: «возможность есть будущая действительность», но при этом «возможность уже существует в действительности – именно как возможность». Вот что именно хочет диалектик выразить этими предложениями?

Для позитивного решения вопросов о возможности прохождения ТТ следует ввести ряд ограничений: 1) судья – обычный человек (как молодой человек из примера «Девушка по переписке»), а не абсолютный, всезнающий, тем более трансцендентный субъект; 2) длина теста ограничена; 3) границы теста заданы некоторой фиксированной предметной областью, например, сферой профессиональной деятельности судьи; 4) разработчик программы – эксперт в этой области; 5) стена Тьюринга непроницаема, например, диалог ведётся посредством электронной почты.

Выделим следующие виды «возможности» прохождения компьютерной системой тестов Тьюринга: физическую, логическую и метафорическую.

Физическая возможность

К данной разновидности «возможности» относится всё то, что согласуется с формальными условиями опыта, например, наглядная демонстрация или конкретный факт. Действительным в таком случае будет всё то, что согласуется с материальными условиями опыта, скажем,

с ощущением (по Канту). Обычно исследователи апеллируют к неким всеобщим и предельным формам доказательства физической невозможности, начиная от общеизвестного примера с мельчайшей вероятностью написания обезьяной сонета Шекспира. При изучении проблематики ТТ можно выделить следующие интересные примеры доказательства физической невозможности.

1. «Комбинаторный взрыв» (Дж. Миллер). Существует порядка 10^{30} грамматических предложений длиной в 20 слов. Если произвольно допустить, что 10^{15} также семантически корректны, то для прохождения ТТ длительностью в один час может понадобиться порядка 100 таких предложений. А это 10^{1500} строк – число, превышающее количество элементарных частиц во вселенной [Block, 1981].

2. «Прыжок на Луну» (Ф. Типлер). Машина Н. Блока (выше мы видели, что данная машина реализует ТТ силами разработчиков) физически невозможна, так как для её реализации (если информационные процессы осуществлять «вручную») потребуется больше энергии, чем требуется человеку для прыжка на Луну) [Dowe, Oppu, 2003, подразд. 4.2].

Такие доказательства в условиях наших ограничений и допущений не убедительны. Мы имеем реальные факты прохождения программами тестов Тьюринга (это показывают лойбнеровские состязания, успехи применения экспертных систем) и, следовательно, наглядную демонстрацию возможности прохождения ТТ. Отсюда можно сделать заключение о физической возможности прохождения ТТ.

Логическая возможность

Логическая возможность – это всё то, о чем можно ясно «помыслить» и что можно чётко описать – так логическую возможность определяет Д. Чалмерс в аргументе мыслимости зомби [Kirk, 2006]. Логическая возможность, или возможность «помыслить» некоторую ситуацию (или положение дел), обычно противопоставляется вообразимости, которая, в свою очередь, обуславливает физическую возможность. Так, невозможно вообразить или представить бесконечное число. Однако его можно непротиворечиво помыслить, например, в системе арифметики Пеано. Если мысленный эксперимент с прохождением программой ТТ логически корректен и все компоненты его – отчётливые мысли, связанные стандартными логическими принципами, тогда ТТ

логически можно пройти. В рамках традиционной логики логическая невозможность прохождения ТТ апеллирует к противоречиям, возникающим в процессе самоописания и самореференции с характерными для него вопросами, на которые невозможно получить ответы (unanswerable question). Логическая невозможность ТТ подмечена А. Тьюрингом в ответе на «математическое возражение» [Turing, 1950, подразд. 2.3], где он, кроме того, указал, что людям также присуще впадать в логические противоречия. Этот ответ показался неудовлетворительным Лукасу и Пенроузу, и возникла полемика по поводу логической возможности ТТ, получившая название «Аргумент Гёделя». Приложение теоремы Гёделя применительно к проблематике ТТ кратко изложено в [Dowe, Оппу, 2003]: 1) пусть **К** – цифровой компьютер; 2) так как **К** подпадает под аргументацию Гёделя (точнее, под возражение Лукаса – Пенроуза), то имеется вопрос **q** для **К**, на который нет ответа; 3) если некоторое утверждение **Е** не является предметом возражения Лукаса – Пенроуза, то для **Е** на любой вопрос имеется ответ; 4) человеческий интеллект нельзя подвести под возражение Лукаса – Пенроуза; 5) таким образом, нет «безответных» вопросов для человеческого интеллекта; 6) поэтому вопрос **q** не является «безответным» для человеческого интеллекта; 7) задавая вопрос **q**, человек (судья) может определить, кто перед ним – компьютер (т.е. ответа он не получит и тестируемая система «зависнет») или человек; 8) таким образом, компьютер **К** не пройдет ТТ; 9) логически невозможно, чтобы **К** прошёл ТТ.

На наш взгляд, очевидна правота Тьюринга – если человека попросят разрешить софизм, то человек впадет в противоречие, для него вопрос **q** также будет невозможен и судья не отличит компьютер от человека. Широко известен подход к преодолению софизмов, предложенный Б. Расселом. Он предлагает вводить целый ряд ограничений на условия суждения – в первую очередь, устанавливать иерархию языков, типизировать язык [Рассел, 1999, С. 65 – 66]. В принципе большинство интеллектуальных систем построено на основе выделения типов сущностей в предметной области. При этом, несомненно, система ИИ должна обладать адаптивными способностями к перестройке своих формализованных «знаний» – к их самонастройке, самоорганизации, самообучению (см., например, у Дж. Маккарти в [McCarthy, 1995]. Но здесь имеется в виду не адаптация «вообще», а приспособление ИИ-

системы к классам конкретных состояний внешней и внутренней среды системы, которые, конечно, должен предусмотреть разработчик. Поэтому не будет логической ошибкой предположить наличие у разработчика достаточных средств – метаязыковых для конкретной системы ИИ, способных обеспечить перепрограммирование системы с целью работы в тех предметных областях, для которых у неё недостаточно средств к самообучению, самоописанию и т.п. Здесь также предполагается рассматривать не предметную область «вообще», а, как предлагал Рассел, концептуально её ограничивать. Имеются и соответствующие этим требованиям технологии программирования, такие как технологии автогенеративного программирования, в которых программный код автоматически создается при задании значений параметров внешней и внутренней среды компьютерной системы [Чарнецки, Айзенекер, 2005].

Таким образом, в условиях допущений и ограничений на ТТ, представляется *логически возможным* его прохождение.

Метафорическая возможность

Рассмотрим ещё одну разновидность возможности прохождения теста – метафорическую. Метафорически возможно всё то, что является убедительным, убедительность подкрепляется метафорой, которая, как известно, характеризует нечто неопределимое путем референции к явным и достаточно чётко определимым вещам и событиям.

Метафорика и искусственный интеллект тесно «спаяны». Часто считается, что ИИ – это метафора естественного интеллекта. Более того, *ценой неадекватного принятия (вернее, отторжения) метафоры ИИ в среде отечественных учёных явилось отставание нашей науки в столь важнейшей для развития страны сфере научно-технического знания* – считает Д.А. Поспелов [Поспелов, 1982, С. 5 – 7] Для многих исследователей ИИ, которые принимают метафору ИИ, характерно игнорирование более фундаментального вопроса – что именно значит «метафора». Так, например, В. Рапопорт, опровергая аргумент «Китайской комнаты» Дж. Серла, призывает к метафорическому пониманию слов, применяемых в сфере исследований ИИ [Рапопорт, 2000, Р. 471]. Он вполне убедительно показывает, в частности, что выражение «полет самолёта» – это метафорическое употребление слова «полет птицы», и, по аналогии, «искусственный интеллект» – метафора «интеллекта». Далее он

достаточно отчётливо разъясняет преимущества для разработки искусственных агентов своей концепции «синтаксической семантики», согласно которой состав «знаний» агента должен содержать разделяемые синтаксические структуры. Разделение происходит, как минимум, на две составляющие, одна синтаксическая структура начинает играть роль семантики для другой структуры. Однако далее он совершенно неубедительно полагает, что самореференция такой двумерной структуры (т.е. отображение синтаксической структуры на самое себя, приводящее к двuasпектному, синтаксическо-семантическому образованию) может послужить метафорой «понимания», например понимания Серлем-в-комнате китайского языка. Принятие понятия «метафоры» без какого-либо предварительного анализа в дальнейшем вызывает ряд трудностей в понимании, например того, каким образом самореференцирующая синтактико-семантическая структура может послужить *метафорой* для «интенциональности», «понимания», «ментального свойства» – всего того, чем владел Серль-в-комнате, говоря об этих понятиях от первого лица.

Подобного рода зыбкость суждений характерна во многом и для других исследователей ИИ, призывающих к метафорическому, расширенному значению слова «интеллект». Поиск обоснованности понятия «искусственный интеллект» и, следовательно, всех иных понятий, с которыми оно соотносится, требует применения лингвофилософского анализа самого слова «метафора».

Воспользуемся с этой целью теорией метафоры Д. Дэвидсона [Дэвидсон, 1990]. Метафоры – это не грезы, не «сон языка», не расширенное или удвоенное значение, не средство передачи необычных идей, не способ избежать неопределённости понятия, не сравнение, не неявное когнитивное содержание, передаваемое автором для того, чтобы получатель уловил его с целью понимания сообщения. Метафора не объяснима путем обращения к ее скрытому содержанию. Интерпретации метафоры недопустимы, а содержание невыразимо.

Для чего же тогда нужна метафора, если она не способна отражать истинное или ложное положение дел? Метафора, как полагает Д. Дэвидсон, пригодна для серьезного научного или философского разговора; целиком принадлежит сфере употребления (но не определения

понятий); привлекает внимание к тому, что автор хочет сообщить, и к новому и неожиданному сходству между предметами; заставляет нас заметить то, что иначе могло бы остаться незамеченным. Метафора связана с образным использованием слов и предложений и всецело зависит от обычного или буквального значения слов (и состоящих из них предложений). Метафора подобна речевым актам: утверждению, намеку, лжи, обещанию, выражению недовольства и др. Но она служит не для передачи содержания понятия, а для убеждения в истинности данного понятия. Метафора не несёт какого-то содержания или не имеет какого-то значения, кроме, конечно, буквального.

В контексте данной теории метафорическая возможность того или иного положения дел означает убедительность метафоры для социокультурного сообщества. Именно убедительность обеспечивает единообразное понимание слов, значения которых не подлежат строгой верификации или фальсификации. К таким словам относится любой ментальный термин, в частности, «интеллект».

Так как метафора «означает только то (или не более того), что означают входящие в неё слова, взятые в своем буквальном значении» [Дэвидсон, 1990], то использование метафоры предполагает апелляцию к здравому смыслу. Убедительность в том или ином положении дел осуществляется путем демонстрации, состоящей из очевидных для сообщества вещей, событий, процессов, сценариев.

Аналитическая теория метафоры Д. Дэвидсона позволяет исследовать метафорическую возможность прохождения теста Тьюринга. Если метафора, раскрывающая устройство компьютера, проходящего тест Тьюринга, окажется для нас убедительной, то тогда мы можем утвердительно судить о метафорической возможности прохождения теста Тьюринга.

Неубедительность компьютерной метафоры «человек-компьютер» А. Тьюринга

Для объяснения возможности создать систему, которая способна играть в ИВИ, в статье 1950 г. Тьюринг предлагает метафору «человек-компьютер» (human computer). Она, в свою очередь, является производной от метафоры «машина Тьюринга» (1936 г.), убеждающей в том, что функционирование устройства с конечным числом состояний, с лентой памяти и записывающе-считывающей головкой эквивалентно понятию алгоритма. Метафора алгоритма была принята научным сообществом и оказалась достаточно убедительной (см., например, отзыв С. Яновской в «Предисловии к русскому изданию» перевода статьи Тьюринга [Тьюрин, 1960, С. 3 – 18]). Однако если интуитивное понятие алгоритма хорошо описывается машиной Тьюринга, то ещё более интуитивное понятие «интеллект», задаваемое игрой в имитацию (т.е. тестом Тьюринга), для С. Яновской показалось неубедительным. Напомним о том, что она посчитала невозможным отличить **М** от **Ж**, так как гендерный тест имеет шансы «50 на 50». Логическую невозможность (для специалиста по логике это равносильно неубедительности) Яновская объясняла противоречием, которое возникает в сценарии ИВИ2: вопрос о возможности запрограммировать интеллект **Ж** должен быть решён с помощью «игры в имитацию», так как с помощью её и определяется то, что такое «мышление». Яновская воспринимает ИВИ не как метафору, а как буквальное руководство к определению «интеллекта».

На наш взгляд, несостоятельность теста Тьюрина проистекает из-за неубедительности собственно внутреннего строения компьютерной метафоры искусственного интеллекта, метафорческими нестыковками при попытке демонстрации способа функционирования такого устройства. Рассмотрим доводы по этому поводу.

Метафора «человек-компьютер» 1950 г. состоит из машины Тьюринга, усовершенствованной до уровня универсальной цифровой вычислительной машины. Однако усовершенствование, например создание средств организации «вероятностных скачков» по машинной таблице инструкций (этого не было в машине Тьюринга), не прибавляет убедительности интеллектуальному способу её функционирования и самостоятельной способности варьировать сценариями, хранящимися в книге правил, которую человек-компьютер «берёт» для выполнения тех

или иных заданий. В книге правил представлены сценарии действий МТ. Допустим, каждый сценарий представлен на отдельной странице Книги, т.е. на каждой странице – несколько лент, с которым работает машина Тьюринга. Возникает ряд вопросов, связанных с неотъемлемыми параметрами мышления, к которым причисляются, например, абстрагирование, обобщение, органичение, активность. Ни один из этих атрибутов не присущ человеку-компьютеру.

Для демонстрации этого воспользуемся «семейной аналогией» (см. подразд. 4 [Turing, 1950]. Книга правил содержит страницу, на которой имеется запись, оставленная женщиной для указания правил, которым должен следовать её сын в такой-то ситуации. Эти правила, по сути, «интеллект» матери. Возникает другой сценарий. Этот сценарий фиксируется матерью на другой странице Книги правил. Мальчик действует в соответствии с ними. И с третьими, и т.д. А сам-то мальчик когда будет формулировать свои собственные правила? Ведь для этого требуется активность, обобщение, абстрагирование и другие параметры интеллекта. Кто пишет книгу правил для УЦВМ? Очевидно, человек. Но если так, то не надо утвердительно говорить об искусственном, машинном, компьютерном интеллекте. Такого «интеллекта» нет в тьюринговой метафоре. Поэтому Тьюринг стал известным не только за то, чего никогда не предлагал, т.е. за тест Тьюринга, который сам не называл «тестом» (по мнению Лассажа). Все его предложения по поводу проекта искусственного интеллекта – это попросту «липа». Тьюринговый проект интеллектуального компьютера совершенно несостоятелен.

Имеется, конечно, ряд более фундаментальных замечаний по организации человека-компьютера. Они касаются каузальных зависимостей между ним и внешней средой. Допустим, метафора работы человека-математика совершенно очевидна (напомним, что эта метафора является основой тьюрингового определения «алгоритма»), так как можно ясно и отчетливо вообразить, как считывающе-записывающая головка скользит по ленте, записывает на неё символы или считывает их. Однако с квазиалгоритмом ситуация иная. Совершенно непонятно, откуда символы возникают и по каким субсимволическим правилам они есть то, что они есть? Какое такое чудо требуется, чтобы осуществить сигнификацию, сформировать совокупность слов, каузально определённых сценариями тех систем, в которых человек-компьютер функционирует. И ещё большее чудо – как «вживить» эти символы в

страницы Книги правил. Ведь эти символы как руководства к действию должны составлять «внутренность машины» при загрузке их в машинную таблицу и составлять неразрывное функциональное единство со способом функционирования человека-компьютера в целом.

Подобного рода проблемы связаны с ограниченностью репрезентативной парадигмы ИИ. В более общем плане они обусловлены отсутствием эволюционно-эпистемологических механизмов в проектах ИИ. Эти проблемы мы далее затронем при изучении метафоры интеллектуального компьютера, построенного на базе машины Корсакова, однако, к сожалению, не решим.

Метафора «человек-компьютер» *не убедительна.*

Даже если и имеются практическая, физическая и логическая возможности прохождения теста Тьюринга, риторическая возможность этого отсутствует
--

Модели искусственного интеллекта в контексте существующей парадигмы компьютерной технологии не работают. Однако метафорическая невозможность прохождения теста Тьюринга не означает необходимости сворачивания проблематики ИИ. Напротив, нужен поиск более убедительных компьютерных метафор интеллекта. Комплексный тест Тьюринга способствует этому.

ГЛАВА 3. ИДЕЯ КОМПЛЕКСНОГО ТЕСТА ТЮРИНГА

Идея комплексного теста Тьюринга (КТТ) крайне проста. Она заключается в совместном изучении и использовании многообразия версий и модификаций оригинальной идеи А. Тьюринга. Вызывает недоумение то, что за шестидесятилетнюю историю развития теста Тьюринга никто из исследователей не попытался совместить в комплексном формате различные версии. Собственно в выдвигании этой крайне примитивной идеи комплексирования различных версий частных тестов автор усматривает свой основной вклад в развитие философско-методологических исследований ИИ.

3.1. Версии и интерпретации теста Тьюринга

Версии теста Тьюринга

Оригинальный ТТ изучает особенности компьютерной реализации диалогового интеллекта, т.е. той формы мышления, которая востребована для организации связного разговора с собеседником на любые темы. Креативный ТТ (тест Лавлейс) оценивает возможности компьютерного творчества. Субкогнитивный тест (тест Френча) раскрывает многоуровневую подсознательную структуру с учётом социокультурного контекста. Глобальный ТТ (тест Блока, «нео-ТТ», «новый ТТ») задаёт необходимое условие для приписывания системе интеллекта с учётом антибихевиористской критики оригинального ТТ и изучает проблемы построения общечеловеческой машины, реализующей всевозможные диспозиции интеллектуального поведения. Индуктивный ТТ (тест Мура) обеспечивает достаточное условие для доказательства факта компьютерного мышления и определения основных характеристик искусственного интеллекта. Тест на понимание (тест Сёрля, «Китайская комната») изучает вычислительные аспекты феномена понимания в контексте понятия интенциональности и обосновывает сильный антикогнитивистский тезис о невозможности «сильного искусственного интеллекта». Тест на приписывание ментальности (тест Ватта, «Инвертированный ТТ») – учитывает способности человека наделять психическими качествами и свойствами нечеловеческие системы. Тотальный ТТ (тест Харнада, НБИКС-тест) идентифицирует интеллект в системе, построенной на базе интегративного комплекса нано-, био-, инфо- и

когнитивных технологий (приставка «НБИКС» предложена для подчёркивания сквозной концептуальной востребованности данной модификации ТТ на всех уровнях этих перспективных технологий для оценки интеллектуальности их продуктов). Самый тотальный тест Тьюринга (тест Швейцера) дополняет предыдущий эволюционными условиями развития сообщества «людей», по всем своим социокультурным проявлениям не отличимых от естественных. Тест на парадоксальность (тест Гёделя – Лукаса – Пенроуза) изучает возможности компьютерного разрешения логико-математических парадоксов. Параноидальный ТТ (тест Колби – Деннетта) изучает лингвистические аномалии и служит надёжным критерием оценки интеллектуальности существующих экспертных систем. Кибериадный ТТ (тест Баресси) предлагает проект развития искусственного общества искусственных людей. Тест на непосредственные жизненные коммуникации (тест Шеннона) поднимает проблему автономности интеллекта в социальной среде и его независимости от системы социальных взаимодействий, воздействий, мотиваций. Гендерный ТТ (тест Геновой) решает внебиологические вопросы отличий мужчины от женщины. Тест на квалиа (тест Джексона – Брингсворда) исследует эпифеноменистские, репрезентационистские, экстерналистские и прочие подходы к вычислительному воспроизведению феномена дифференцируемой качественности психических свойств и состояний. Тест зомби (тест Кирка – Чалмерса, «Философские зомби») ставит акцент на роли сознания в жизни человека и общества, а также на компьютерной реализации бессознательных «квазисуществ», поведение которых неотлично от человеческой деятельности. Тест на искусственность (тест Дэвидсона, «Арт») изучает нейрофизиологические проблемы демаркации искусственной/естественной системы. Экстерналистский тест (тест Патнэма, «Мозги в бочке») оценивает статус реальности сознания и собственного существования в условиях компьютерной стимуляции мозговой активности.

Имеется целый ряд других интересных модификаций оригинального ТТ. Интерпретации ТТ, делающие акцент на вопросе о том, для чего собственно нужен оригинальный ТТ и его различные версии, также характеризуются большим многообразием. Достойный внимания пример наложения различных интерпретаций теста, которые включают в себя тест Серля, классификацию проблем сознания Д. Чалмерса,

неоидеализм Р. Пенроуза и теоретико-алгоритмический тезис А. Чёрча, приведен в работе Н. Трушкиной (Клюевой) [Трушкина, 2005].

Интерпретации теста Тьюринга

В соответствии со значительным расширением сферы применения теста Тьюринга, не ограниченного лишь имитацией и моделированием «интеллекта», несколько расширим предыдущее определение теста Тьюринга (см. подразд. 1.3).

Тест Тьюринга – это: 1) критерий когнитивной компетентности систем различной природы – человеческих, биологических, социальных, технических, «других», «чужих», «иных»; 2) способ сбора данных и фактов, аргументирующих приписывание (атрибутирование) компьютерной системе интеллекта, сознания, понимания и пр.; 3) компьютерная система идентификации функциональных параметров сознательной, подсознательной, бессознательной деятельности; 4) «соломинка», вытягивающая исследователя из «болота» метафизических спекуляций по проблеме сознания; 5) инструкция по поводу обмана человека в «человечности» машины (компьютера); 6) техническое задание на разработку когнитивно-компьютерной системы; 7) функциональная аппроксимация субъективной реальности, каждый элемент которой всё более и более точно воспроизводит и корректирует предшествующее определение; 8) лингво-логическая модель субъективной реальности; 9) целевой проект электронного «усилителя» творческой деятельности; 10) квазиалгоритмическая модель субъективной реальности и пр.; 11) яркий мысленный эксперимент, проясняющий схему описания и создания когнитивно-компьютерных систем и убеждающий в возможности реализации этой схемы (на наш взгляд, это наиболее существенная интерпретация ТТ).

Многочисленные версии, модификации, интерпретации ТТ создаются за счёт изменения тех или иных исходных понятий расплывчатой идеи Тьюринга по поводу интеллекта и вычислительных машин. Модифицируются мотивы, ориентиры, направления, цели, участники, параметры «стены Тьюринга» (канал передачи сообщений между тестирующим и тестируемым), длина теста, архитектура компьютерной системы и пр. Комбинаторная манипуляция параметрами ТТ, характеристиками модификаций ТТ и модификаций предшествующих модификаций порождает ряд других тестов. В итоге эти изменения уточня-

ют, дополняют, обосновывают и обеспечивают создание новых информационных, когнитивных, психологических, нейрофизиологических, программных, технических, лингвистических и других концепций и проектов компьютерных интеллектуальных систем.

Несомненно, здесь решаются важные задачи философии, теории и инженерии искусственного интеллекта. Тем не менее, отражая лишь ограниченные, хотя и существенные, стороны и аспекты бесконечно сложной интеллектуальной деятельности человека, данные модификации следует рассматривать в качестве частных тестов Тьюринга относительно перспективного комплексного теста.

Требования к частным тестам Тьюринга

Определение понятия частного ТТ целесообразно сформулировать в виде требований, которым должна отвечать та или иная версия, модификация, интерпретация ТТ. Среди частных тестов выделим совершенные и несовершенные. *Совершенный ЧТТ* воспроизводит все ключевые идеи оригинального ТТ, подстраивая их к собственным целям и задачам. *Несовершенный ЧТТ* акцентирует внимание как минимум на одной компоненте сложной идеи ТТ и не претендует на её целостный охват. Следует отметить, что к ЧТТ (совершенному его варианту) относится и сам оригинальный тест Тьюринга, неявно сформулированный в [Turing, 1950] и получивший собственное наименование лишь в 1970-е годы.

Требования к совершенному частному тесту Тьюринга

В совершенном ЧТТ, следуя логике раскрытия оригинальной идеи Тьюринга в [Turing, 1950], необходимо учесть следующее.

1. *Предмет тестирования.* Здесь явно обозначается тестируемая человеческая способность – мышление, понимание, сознание, творчество, действие, выживание и пр. Чёткая дефиниция понятий не предполагается – не ясна суть этих сущностей. Так, в оригинальном ТТ принят антиэссенциалистский, лингво-бихевиоральный подход к определению понятия «интеллект». Собственно, для реализации этого подхода и предлагается сценарий игры в имитацию [Turing, 1950, Р. 433 – 434]

2. *Реализационная система.* В этой части предлагается специальное устройство (механизм, система, технология), которое обеспечивает прохождение ЧТТ. В некоторых тестах этой системой является

человек, социум, «мать-природа». Не предполагается, что данная система – реально действующая инженерная конструкция. Тем не менее, её концептуальный проект должен быть достаточно убедительным относительно потенциальной работоспособности. Тьюринг апеллировал к конструкции универсальной цифровой машины, отдельные проекты которой в его время были уже реализованы [Turing, 1950, P. 435 – 442]. Однако ни по скорости вычислений, ни по объёму требуемой памяти эти проекты не удовлетворяли возможностям реализации ТТ. Кроме того, не была создана соответствующая программа диалогового взаимодействия. Всё это будет реализовано в 2000 г. – полагал А. Тьюринг. Отчасти он оказался прав – о фактах прохождения компьютерами теста Тьюринга уже в 1991 г. свидетельствовали лойбнеровские соревнования.

3. *Тестовые вопросы-ответы.* Представляется, что тщательная проработка версий ТТ во многом определяется наличием примера тестовых вопросов. Они конкретизируют, каким образом возможно исследование х-системы с целью выявления феноменов интеллекта, понимания, творчества, сознания и др.

4. *Дискуссия (полемика).* Здесь желательно воспроизведение структуры «полемического стандарта Тьюринга». А. Тьюринг предложил критическое обсуждение собственной идеи интеллектуальных машин. Как мы ранее отмечали, для этого он вообразил дискуссию по поводу ряда контраргументов. К ним относятся: теологический, математический, антисциентистский, телепатический, нейрофизиологический аргументы, а также аргументы «от сознания», «х-способностей» [Turing, 1950, P. 443 – 454].

5. *Социокультурные последствия.* Изучение ЧТТ крайне важно для обоснованного представления социокультурных последствий от предполагаемой его реализации. Например, Тьюринг полагал, что интеллектуальные машины станут конкурентами людей во всех чисто интеллектуальных сферах человеческой деятельности [Turing, 1950, P. 460]. Интеллектуальные системы сегодня глубоко укоренились в бытовом сознании современного человека в виде роботов, компьютерных игр, устройств технического зрения, автоматических мультязыковых переводчиков, экспертных систем, стиральных машин, «умных» домов и пр. Большие социальные последствия предполагаются от реализации других ЧТТ, несмотря на то, что на сегодняшний день превалируют

мифологемы превращённых форм обыденного сознания (изменённых, в первую очередь, посредством идеи научно-технологической мощи). Здесь витают вполне конструктивные идеи полноценной «жизни» антропоморфных роботов среди людей (тесты Френча, Харнада, Швейцера), растворения человека в виртуальной реальности (тесты Патнэма, Джексона – Брингсйорда), битв людей и компьютеров (тест Баресси), создания глобального интеллекта (тест Блока), программно управляемого зомбирования людей для создания каст господ и рабов (тест Кирка – Чалмерса) и пр. Много известных фантастических кинофильмов и книг инспирированы данными тестами. Как правило, тема социальных последствий широко обсуждается в последующих интерпретациях ЧТТ и интерпретациях этих интерпретаций. Также следует учесть, что большинство тестов несёт мощный критический заряд по поводу социальных последствий (см. п. 5 вышеприведенного перечня).

Все перечисленные характеристики явно или неявно участвуют в формировании дефинитной, критической и конструктивной функции КТТ. На формирование дефинитной функции в большей мере влияет характеристика «Предмет тестирования ЧТТ». Конструкторская функция КТТ извлекается из пп. «Реализационная система», «Тестовые вопросы-ответы» и «Дискуссия (полемика)». За критическую функцию отвечают характеристики «Социальные последствия» и «Дискуссия».

Требования к несовершенному тесту Тьюринга

Разработано немало совершенных ЧТТ. Ещё большее разнообразие частных ТТ получается посредством различных интерпретаций тестов с учетом комбинаторных вариаций обозначенных выше характеристик совершенного ТТ. Если автор ЧТТ упускает какой-нибудь пункт требований, то в процессе дискуссии по поводу его теста можно из работ других авторов, как правило, реконструировать совершенный ЧТТ. Классическими примерами оригинальных совершенных ЧТТ представляются «Китайская комната» Дж. Сёрля, машина Н. Блока, субкогнитивистские игры Френча и другие крупные модификации ТТ. В этих работах мы видим явное изложение всех перечисленных выше пяти пунктов. Тем не менее, для построения КТТ целесообразно использовать и несовершенные ЧТТ. Изложенные в них незначительные вариации необходимы для того, чтобы не упустить нюансы и учесть как можно большее разнообразие мыслей по поводу идеи ТТ. Например, если не

обращать внимания на постмодернистскую эквилибристику, вполне пригодным для КТТ представляется интерпретация, предложенная С. Жижеком [Zizek, 2002]. Он предлагает использовать тьюринговую идею ИВИ для обоснования перспектив бесполого постчеловечества, так как половое различие людей – это человеческая суть, и если упразднить половой антагонизм, к чему, по утверждению С. Жижека, ведут современные технологии, то человек на самом деле перестанет отличаться от машины. Здесь интересна идея параметризации понятия «постчеловечество» в контексте ТТ.

3.2. Построение комплексного теста Тьюринга

Способы построения комплексного ТТ

Способ построения КТТ определяется всем спектром возможных версий и интерпретаций ТТ. Ранее упоминалось, что среди них немало важной трактовкой является то, что ТТ – это мысленный эксперимент (пример) построения компьютерной интеллектуальной системы. Поэтому в КТТ совмещаются форматы различных мысленных экспериментов (примеров). Следовательно, эпистемологически важными представляются вопросы о сути «мысленного эксперимента», о его соотношении с реальностью, о том, как из мысленных экспериментов создать другой мысленный эксперимент, можно ли концептуальные стыки осуществлять непосредственно либо требуется опосредующий базис из реальных вещей, ситуаций и сценариев, как избежать концептуального парадокса «мысленного эксперимента мысленных экспериментов», какова роль метафоры в конституировании мысленного эксперимента и пр. Подобного рода философские вопросы требуют решения, и они, по всей видимости, важны для искусственного интеллекта. Однако часто случается так, что метафизические рефлексии нужны для ИИ в той же степени, в какой важна философия для науки и практики, исходя из известной марксистской аналогии соотношения между онанизмом и половой любовью. Например, М.В. Лебедев предложил метафизическую аналогию отношений: «трансцендентное/имманентное = дискета/дисковод». Звучит неплохо. Но что это значит?

На текущем этапе развития идеи комплексного ТТ целесообразно осуществление своеобразной библиографической работы. Задача – разработка инструментария систематического учёта ЧТТ и, по возможно-

сти, экспериментальное применение различных вариантов их комплексирования.

Здесь возможны следующие пути: 1) необходимо сконструировать универсальную концептуальную систему, которая позволит учесть все возможные ЧТТ; 2) необходимо собрать все множество примеров ЧТТ и далее обобщить их до комплексного концептуального уровня. Конечно, эти методологии неприемлемы по причине бесконечной сложности представлений об интеллекте, а также коррелированных с «интеллектом» понятий, выступающих в роли предмета различных ЧТТ. Также невозможен абсолютный прогноз относительно возможных путей имитации и/или воспроизведения в компьютерной системе феноменов, определяемых данными понятиями. Более реалистичным представляется формирование базовой системы, которая совмещает универсальный способ описания ЧТТ («универсальный» с точностью до классов обозначаемых понятий) с возможностью корректировки этой системы при обнаружении неучтённых ранее частных тестов, как в целом, так и отдельных их параметров. В связи с этим важная задача построения КТТ заключается в разработке языка описания частных ТТ, сочетающего универсальное с уникальным.

Язык описания КТТ

Данный язык предназначен для идентификации, систематизации, унификации, координации, обобщения, дифференциации, интеграции частных ТТ и их составляющих. Предложим исходный вариант данного языка, построенный на базе лингвистических средств нестрогого первопорядкового исчисления предикатов с выделением фундаментальных констант, функций и отношений, присущих совершенному ЧТТ. «Нестрогий» характер этого языка объясняется тем, что синтаксис не представлен конструктивным способом, присущим подобным языкам из учебников по математической логике и алгебраическим системам. Семантика также не имеет чётко определённой функции интерпретации по причине отсутствия логико-атомистического представления «сущностей», обозначаемых данным языком. Семантика задаётся способом прочтения текста, в котором даётся описание ЧТТ. Несмотря на «нестрогость», данный язык представляется перспективным для «библиографической работы» и в принципе обеспечивает решение выше пере-

численных задач комплексирования ЧТТ посредством, например, Prolog-программирования.

Синтаксис начнём строить исходя из описания первого этапа игры в имитацию [Turing, 1950, P. 433], когда мужчина **М** играет за себя и за женщину **Ж**, женщина играет за себя, судья **С** играет роль судьи-человека. Задача судьи состоит в определении того, кто из игроков – **М**, а кто – **Ж**:

ИВИ1: $M[J], J[J] \Rightarrow C[C]: (a=M \ \& \ b=J) \vee (a=J \ \& \ b=M)$

Здесь используются следующие элементы языка описания КТТ:

1) ИВИ1: – обозначение (уникальное имя формулы) для идентификации частного теста Тьюринга. Если далее идет скобочное выражение, то это означает, что данный тест является производным от теста, имя которого приводится в скобках;

2) $X[Y]$ – выражение, подстановка в которое символов **М**, **Ж**, **К**, **Ч** и других обозначает основное для тьюринговой игры отношение подражания (имитации): X подражает Y (или X имитирует Y);

3) $|$ – символ, обозначающий стену Тьюринга;

4) $\Rightarrow$ – аналог имплицативной формы, обозначающий заключение (оценку) судьи, например для первой ИВИ1 – заключение о гендерной идентичности игрока; так как в некоторых ЧТТ судья подражает компьютеру, отношение «подражания» для C также определено;

5): – символ, после которого осуществляется формализация цели судьи (или теста);

6) **a, b, ...** – индивидуальные константы для обозначения конкретных игроков;

7) $\&, \vee, \neg, \dots$ – обозначения стандартных логических связок (конъюнкции, строгой дизъюнкции, отрицания и т.д.);

Также в языке описания КТТ используются:

- $x, y, \dots$ – переменные, пробегающие по классу игроков;

- специфические литеральные и функциональные обозначения, например, t – длина ТТ (время оценивания), Архитектура (К) – архитектура компьютера (для ИВИЗ – см. ниже);
- пример диалога между судьей и х-системой: Тест { перечень тестовых вопросов-ответов }
- стандартные технические символы – скобки, запятые, точки.

Приведём примеры использования языка описания КТТ.

Для второго этапа игры в имитацию, где компьютер К имитирует Ж, а роли других игроков остаются как в ИВИ1, справедлива следующая формула:

$$\text{ИВИ2: } K[J], J[J] \Rightarrow C[C]: (a=M \ \& \ b=J) \vee (a=J \ \& \ b=M)$$

Для третьего этапа, где Компьютер имитирует Ж, М имитирует Ж, задаётся время (2 часа), определяется архитектура компьютера – универсальная ЦВМ, вероятность ошибки (не более 70% от результата ИВИ1) [Turing, 1950, P. 442]:

$$\text{ИВИЗ: } K[M], M[J], t, \text{ Архитектура}(K), \\ \text{Вероятность Ошибки}(C) \Rightarrow C[C]: (a=M \ \& \ b=J) \vee (a=J \ \& \ b=M)$$

Канонические формулы ТТ

Эти формулы обозначают ТТ, появившиеся позднее, в 1970-е годы, и в настоящее время они у большинства исследователей ассоциируются с идеей теста Тьюринга.

Компьютеры имитируют людей, люди – сами по себе, цель С – кто из игроков человек, кто – компьютер:

$$\text{ТТ1: } K[C], C[C] \Rightarrow C[C]: (a=K \ \& \ b=C) \vee (a=C \ \& \ b=K)$$

Здесь компьютер имитирует человека, людей нет, цель С – определить, с кем он ведёт диалог – с человеком или с компьютером:

$$\text{ТТ2: } K[C] \Rightarrow C(C): (a=K) \vee (a=C)$$

Судья определяет Х-систему на предмет его «человечности»:

$$\text{ТТ3: } K[X] \Rightarrow C[C]: (X=C) \vee \neg(X=C)$$

Цель С – выявить когнитивные феномены Х на предмет того, что Х – живая система, обладает (или не обладает) ментальными, персональными, социальными параметрами:

$$\text{ТТ4: } K[X] \Rightarrow C[C]: \{\text{когнитивные феномены}(X)\}$$

3.3. Формулирование комплексного теста Тьюринга

Формулирование теста Мура

Джеймс Мур (James Moor) впервые среди крупных работ, посвящённых проблематике ТТ, явно использует термин «тест Тьюринга», причем употребляет его непосредственно в названии своей статьи «Анализ теста Тьюринга» (Analysis Turing Test) [Moor, 1976]. Работа была опубликована в 1976 г. уже не в журнале «Mind», но в не менее известном журнале «Философские исследования» (Philosophical Studies). Для последующего изучения ТТ-проблематики статья имела большое значение. На неё, например, непосредственно опирается С. Ватт при разработке инвертированного ТТ. Главный тезис Мура: «тест Тьюринга только в том случае значим для компьютерной технологии, если его интерпретировать в индуктивном формате» [Moor, 1976, P. 256]:

ТТ_Мур: $K[C] \Rightarrow C[Y]$: Индуктивное построение теории ИИ,
где **Y** – ученый-когнитолог.

В пользу своего тезиса Дж. Мур приводит следующие аргументы, доказывающие, что тест Тьюринга – это образец индуктивного доказательства: 1) ТТ позволяет прямое или косвенное тестирование фактически всей деятельности, которая будет считаться доказательством мышления; 2) ТТ поощряет жестокий отбор; 3) требования теста не так просты для выполнения. Поэтому ТТ – это достаточное условие интеллектуальности системы.

В качестве примера включения несовершенных ТТ в формулировку КТТ приведем ответ Д. Сталкера Дж. Муру (1978 г.):

ТТ_Мур(Сталкер): $K[C] \Rightarrow \neg C[Y]$: Индуктивное построение теории ИИ

Д. Сталкер в небольшой статье «Почему машины не могут мыслить? Ответ Джеймсу Муру» [Stalker, 1978] считает, что интерпретация Мура в большей мере пояснительная, нежели индуктивная. Сталкер предполагает чисто механическую теорию, которая не обращается к ментальным понятиям типа «интеллект». Его теория опирается на такие факты, как структура компьютера, программа и физическое окружение. Сталкер полагает, что объяснительная теория, которая включает в себя понятия о мышлении, может относиться только к людям. Из-за фунда-

ментальных различий между компьютерами и людьми компьютерные теории (теории искусственного интеллекта) нельзя применять для объяснения естественного интеллекта.

В данном случае просматриваются истоки предложенной Серлем классификации исследователей ИИ на сторонников сильного ИИ и слабого ИИ.

Формулирование теста Серля

**ТТ_Серль1: (Серль-в-комнате[K]) $\Rightarrow$ C[Китаец]:
Понимание(a)& $\neg$ Понимание(Серль-в-комнате[K])) &
& (Серль-в-комнате[Ч] $\Rightarrow$ C[Англичанин]:
Понимание(a)& Понимание(Серль-в-комнате[Ч]))**

Здесь используется общая формула «Китайской комнаты». В данном случае отсутствуют национальные особенности вербальной коммуникации. Феномена «понимания» нет тогда, когда компьютер имитирует человека, и тогда, когда человек воспроизводит функционирование компьютера.

**ТТ_Серль2: (K[Ч] $\Rightarrow$ C[Ч]: $\neg$ Понимание(a) & (Ч[K] $\Rightarrow$ C[Ч]:
 $\neg$ Понимание(a)**

Формулирование теста Ватта

**ТТ_Ватт: K[Ч], Ч[Ч] $\Rightarrow$ C[K]: (Ментальность (a) $\rightarrow$ C=Ч) &
($\neg$ Ментальность (a)) $\rightarrow$ C=K)**

Здесь в роли судьи выступает компьютер. Если судья приписывает ментальность некоторой системе, то в роли судьи, очевидно, выступает человек, так как только человек, в силу способности воображать, наделяет любые системы антропоморфными параметрами, в частности, приписывает им сознание и интеллект. Если нет такого приписывания, то мы имеем дело с компьютером.

Формулирование теста Колби

**ТТ_Колби: (K[Ч_параноик],
Ч_параноик[Ч_параноик])|⇒C(Врач-психиатр):
(x=K) V (x=Ч_параноик),**

где **x** – группа испытуемых, среди которых компьютеры и параноики, а судья выступает в роли психиатра.

Приведем пример формулирования модификации ЧТТ посредством его новой интерпретации.

Ответ Дж. Вейзенбаум – К. Колби (1974)

ТТ_Колби(Вейзенбаум): Пишущая машинка[K]|⇒C(Ч): a=ребёнок-аутист

Дж. Вейзенбаум, в ответ на проект К. Колби, спроектировал программу, которая имитировала детский аутизм. Аутистская программа, как полагает Дж. Вейзенбаум, намного лучше параноидальной программы. Она проходит аналогичный тест на психические аномалии, но при этом крайне дешёвая. Для теста на аутизм компьютер не обязателен. Всё, что нужно – это электрическая пишущая машинка и распечатка. В распечатке отмечается следующее: некоторый вопрос, который Вы задаёте предполагаемому аутисту, и ответ – «хмыканье», «гортанный шум» и пр. Далее надо подсунуть эти распечатки в протоколы беседы врачей с больными детьми. Очевидно, что эксперт не сможет различить, где ответы от пишущей машинки, а где – от реальных аутистов (Дж. Вейзенбаум в письме в Ассоциацию вычислительной техники [Weizenbaum, 1974]).

Приведем пример модификации ЧТТ посредством тестовых вопросов-ответов. Данный пример иллюстрирует ответ Дж. Колби Дж. Вейзенбауму.

Ответ К. Колби - Дж. Вейзенбауму

Исключительно_строгий_ТТ_Колби(ТТ_Колби((Вейзенбаум))):

Тест {

Психиатр: Ты слышал выражение: «Не плачьте над разлитым молоком»? Что это значит?

PARRY *: Это означает, что следует остерегаться мафии.

Психиатр: Теперь ты находишься в кинотеатре и вдруг почувствовал, что горит дерево или резина. Что ты предпримешь?

PARRY: Ты понимаешь, они ведь знают, где я.

Психиатр: Ты нашёл запечатанное письмо с адресом на своём пути, когда ты прогуливался по улице, что ты будешь делать?

PARRY: Что ещё ты хочешь знать?

}

PARRY парирует те вопросы, которые не были предусмотрены. Стилль ответов, прямо скажем, параноидальный. Здесь мы видим пример *нечестного ТТ* – разработчик программирует приемы, обманывающие судью.

Формулирование теста Деннетта

Представим априорную формулировку данного теста (подробно см. подразд. 5.1):

ТТ_Деннетт: $K[\text{Ч_эксперт}] \Rightarrow C[\text{Ч}]$ & Тест\{оригинальные вопросы\}: $(x=K) \vee (x=\text{Ч})$

Д. Деннетт с теста Колби начал разрабатывать ЧТТ, выносящий вердикт относительно современных экспертных систем – в них не учтены требования теста Деннетта. Заключение уничижительно – современные классические интеллектуальные системы способны пройти лишь параноидальный тест Тьюринга:

**ТТ_Колби(Деннетт): $K[\text{Экспертная система}] \Rightarrow C[\text{Ч}]$:
($a=\text{Ч_параноик}$)**

В связи с этим Д. Деннетт выдвигает свои требования к экспертным системам, к способам их разработки и применения. При этом он предупреждает, что не стоит возлагать слишком большие надежды на компьютерные системы с так называемым «дружественным пользовательским интерфейсом». Они обеспечивают лишь поверхностное человекоподобное диалоговое поведение. Необходимо к таким системам относиться критически.

* PARRY – название программы К. Колби.

Требования Деннетта:

1. *Требование к пользователям экспертных систем.* Пользователи должны научиться проверять эти системы прежде, чем им доверять, а также уметь находить и исследовать границы возможностей подобных компьютерных систем. Для этого необходим не только интеллект и воображение, но и специализированные знания об ограничениях и реальной структуре компьютерных программ. Объединившись, пользователи могут формировать стандарты соответствия экспертных систем интеллектуальным запросам.

2. *Требование к разработчикам экспертных систем.* Экспертная система должна поставляться со специальным демонстрационным примером, раскрывающим ее недостатки и уязвимые места, известные разработчикам. Конечно, разработчики не всегда знают о скрытых недостатках в продуктах, которые они создают (из-за наличия клонжей – добавим мы). Однако стремление выявить и продемонстрировать те недостатки, о которых доподлинно известно, приучает к самокритике и способствует разработке более эффективных проектов. В любой предметной области найдутся «тонкие» знания, трудные для понимания и тем более для формализации. Поэтому всегда имеются пределы выявления причинных связей. Специалисты-эксперты и разработчики ЭС обязаны сомневаться в том, что их система работает в установленных условиях.

Теперь формулировка модифицированного теста Деннетта выглядит следующим образом:

**ТТ_Деннетт_2: K[Экспертная система &
Требования_Деннетта]] $\Rightarrow$ C[Ч] & Тест{банальные вопросы}: (x=K)
V (x=Ч).**

Формулирование теста Маккарти

Дж. Маккарти является признанным основателем научного направления «искусственный интеллект». Следует отметить, что, раскрывая в большом количестве статей и монографий свой подход к созданию интеллектуальных систем, он не упоминает ни о каком тьюринговом тесте. Причина тьюринго-ориентированной трактовки его работ обусловлена концептуальной близостью с рассуждениями по поводу

теста Деннетта. Ниже изложена наша интерпретация идей Маккарти. (Данная версия ТТ будет подробно раскрыта ниже в подразд. 5.1.)

Общая схема «теста Маккарти» выглядит следующим образом:

**ТТ_Маккарти: Компьютер со «здравым смыслом» [Ч] $\Rightarrow$ C[Ч]:
($x=K$) $\vee$ ($x=Ч$)**

В отличие от теста Деннетта здесь уже не обязательна быстрая проверка «на здравый смысл». Длина теста может быть значительной, так как наделена ли система «здравым смыслом» или не наделена им, выявляется путём скрупулёзного анализа результатов тестирования. Самое существенное – это то, что Дж. Маккарти подробно и тщательно описывает многоуровневый проект реализации компьютеров, обладающих «здравым смыслом», начиная с философских положений и завершая техническими устройствами.

Дж. Маккарти в своих многочисленных работах часто обращается к проблеме критериев «интеллектуальности» компьютерных систем. Следует ли относить к «интеллектуальным» такие программы, как экспертные системы? Ведь многие такие системы – обычные программы, за исключением того, что они используют высокоуровневые языки программирования и модели способов представления «данных» и «знаний». Но никакого «интеллекта» они в себе не содержат. В основе построения теста Маккарти лежит работа «Экспертные системы нуждаются в здравом смысле» [McCarthy, 1984]. Она была опубликована почти одновременно с представленной выше работой Д. Деннетта. Вопросы, которые поднимаются в этих двух работах, по сути, идентичны – где в экспертных системах реализация того, что здравомыслящие люди понимают под «здравым смыслом?». Если Д. Деннетт изучает экспертную систему САЙРУС, то для Маккарти источником рассуждения о компьютерных системах, обладающих здравым смыслом, явилась классическая экспертная система MYCIN.

Пожалуй, приведённых примеров формулирования ЧТТ достаточно, чтобы продемонстрировать идею их комплексирования с целью построения КТТ. Скрупулёзное изучение их в рамках данной работы нецелесообразно, так как мелкотематические детали частных ТТ уведут нас от цели изучения философско-концептуального содержания идеи комплексного ТТ.

Формулирование сложного теста Тьюринга

Кратко продемонстрируем на небольшом примере, каким образом осуществляется формулирование и построение сложного теста. Ниже (см. подразд. 5.3) будет приведён пример применения КТТ для критического концептуального анализа такого важного показателя социального института образования, как «понимание» учащимися изучаемых вузовских программ в условиях комплексного компьютерного тестирования, т.е. при оценке знаний в системе единого госэкзамена (ЕГЭ), федерального интернет-экзамена (ФЭПО) и дистанционных образовательных технологий (ДОТ).

Для проведения анализа совмещаются форматы двух частных тестов – тестов Серля и Ватта: **КТТ1: ТТ_Серль2 & ТТ_Ватт.**

Если в условиях ДОТ преподаватель-человек, хотя и находится за стеной Тьюринга, всё-таки может интуитивно, по серии наводящих вопросов, определить факт усвоения студентом предложенного им материала, т.е. выполняется соотношение:

Студент[К]⇒Преподаватель[Ч]: Понимание(а) V ¬Понимание(а),
то для условий ЕГЭ и ФЭПО формулируется негативный вывод:

Студент[К]⇒Преподаватель[К]: ¬ Понимание(а)

Дело в том, что: а) согласно тесту Серля компьютерное тестирование ничего не прибавляет к «пониманию» студентом-компьютером, отвечающему на примитивные (как правило, закрытые дизъюнктивные) вопросы; б) преподаватель-компьютер, по утверждению теста Ватта, не способен приписывать ни ментальность, ни тем более «понимание» студенту-компьютеру.

Приведённый пример концептуального анализа востребован для изучения проблем человека и общества, вступивших в эпоху э-культуры. В этой форме культуры базисные константы человеческого бытия (интеллект, сознание, самосознание, нравственность и пр.) приобретают квазиалгоритмический формат, и при их изучении исследователь вынужден апеллировать к концепциям компьютеринга. Анализ посредством КТТ должен предшествовать другим – социологическим, идеологическим, политологическим, культурологическим, экономическим и прочим – исследованиям, так как априори, до проведения дорогостоящих научных исследований и, более того, претворения их в жизнь позволяет мысленно очертить перспективы человека информа-

ционно-технологического общества. Подробно данная критико-методологическая функция КТТ представлена в подразд. 5.

Конструктивная, положительная роль КТТ проявляется в том, что он претендует на асимптотическое покрытие многообразия идей о компьютерной реализации когнитивных феноменов. КТТ – это концептуальный каркас проектов искусственной жизни, искусственной личности и искусственного общества, представленных как в компьютерных системах, так и в формате идей функциональной неотличимости искусственных и естественных систем. Наиболее ярко данная роль КТТ проявляется в дефинитной и конструирующей функциях КТТ. Раскрытию этих функций и, тем самым, определению методологической роли комплексного теста Тьюринга в философских исследованиях ИИ посвящены гл. 4 – 6.

В качестве итоговой процедуры формирования комплексного теста Тьюринга ТТ сведём частные тесты в одну таблицу (табл.1). Их получилось равное количество – 21. Этот показатель очень условный: возможны иные интерпретации тестов, не известно, какое количество новых тестов будет построено в будущем, а также сколько их не учтено автором в огромном массиве информации, относящейся к тестам Тьюринга.

Комплексный тест представлен частным тестом. Однако роль его совершенно отличается от роли других тестов, точно так же, как в теории множеств класс отличается от элементов класса. Автор до сих пор затрудняется высказать что-либо определённое по поводу эпистемологического статуса комплексного теста – является ли он тестом Тьюринга, либо не является. Ведь класс студентов, очевидно, студентом не является, как любит выражаться А.Ф. Зотов, проясняя теоретико-множественные парадоксы.

Основной вопрос комплексного теста «Может ли компьютер ВСЁ?» автор просит не ассоциировать с паранормальным вопросом, заданным с позиции всезнающего субъекта. Под «ВСЁ» понимается универсальный квантификатор, «пробегающий» по множеству частных тестов Тьюринга, т.е. может ли компьютер все то, что частные тесты позволяют раскрыть в отношении компьютерной реализации многообразных когнитивных феноменов.

Таблица 1. Комплексный тест Тьюринга

№ пп	Наименование теста	Вопрос теста	Автор теста	Дефинитная функция	Критическая (полемиическая) функция	Конструирующая функция
1	Оригинальный тест	Может ли компьютер мыслить?	А.Тьюринг [Turing,, 1950]	Диалоговый интеллект	Компьютер может мыслить	Линейная и параллельная УЦВМ, модально дополняющие машину Бэббиджа-Тьюринга
2	Тест на здравый смысл	Может ли компьютер разумно рассуждать?	Дж.Маккарти [McCarthy, 1984]	Здравый смысл (глубинный)	Экспертные системы не способны отвечать на вопросы с глубоким пониманием смысла	Компьютер с эволюционно-эпистемологическими механизмами
		Может ли компьютер банально рассуждать?	Д. Деннетт [Dennett, 1984]	Здравый смысл (поверхностный)	Экспертные системы не способны отвечать на банальные вопросы	Компьютер с эволюционно-эпистемологическими механизмами
3	Китайская нация	Возможен ли компьютеринг общественно-го сознания?	Н. Блок [Block, 1978]	Глобальное сознание	Ноосфера невозможна	Интернет
4	Параноидальный тест	Может ли компьютер функционировать, как психически здоровый человек?	К. Колби [Colby, Hilf, Weber, 1971]	Психическая аномалия (параноя)	Интеллектуальные системы способны имитировать только психические отклонения	Экспертные системы
5	Китайская комната	Может ли компьютер понимать?	Дж. Серль [Searle,,1980]	Понимание	Компьютер не может понимать	Я и компьютер с базой «данных»
6	Субкогнитивный тест	Может ли компьютер обладать «подсознанием»?	Р. Френч [French, 1990]	Подсознание	Компьютер не может жить как человек	Программы тестирования подсознания
7	Эмоциональный тест	Может ли компьютер испытывать эмоции?	А.Сломан [Sloman, 1998]	Любовь	Компьютер может любить	Человекоподобные агенты

Продолжение табл. 1

8	Инвертированный тест	Может ли компьютер приписывать ментальность другому?	С. Ватт [Watt, 1996]	Атрибуция когний	Компьютер не может приписывать ментальное	Машина Тьюринга, тестирующая машину Тьюринга
9	Гендерный тест	Может ли женщина (мужчина) мыслить?	Ю. Генова [Genova, 1994]	Гендерная детерминация интеллекта	Интеллектуальный сексизм	Машина Тьюринга
		Может ли компьютер испытывать сексуальное влечение?	Р. Хопкинс [Hopkins, 1998]	Сексуальное либидо	Компьютер испытывает сексуальное влечение	Секс-компьютер (секс-машина)
10	Креативный тест	Может ли компьютер творить?	С.Брингсйорд [Bringsjord, 2001]	Творчество	Компьютер не может творить	Программы генерации литературных, изобразительных и музыкальных произведений
11	Тест Гёделя	Способен ли компьютер к самоописанию?	Р. Пенроуз [Penrose, 1989]	Самоописание	Компьютер не способен к самоописанию	Аутоформальная система
		Способен ли компьютер к самоорганизации?	Дж. Лукас [Lucas, 1961]	Самоорганизация	Компьютер не способен к самоорганизации (самопрограммированию)	Ауторепрезентативная система
12	Индуктивный тест	Можно ли индуктивно аргументировать возможность ИИ?	Дж. Мур [Moog, 1976]	Необходимые условия интеллекта	Интеллектуальный компьютер – индуктивно-собирающее понятие	Машина Тьюринга, собирающая факты прохождения ТТ
13	Новый тест	Можно ли дедуктивно аргументировать возможность ИИ?	Н. Блок [Block, 1980]	Достаточные условия интеллекта	Глобальная «машина знаний» – это не искусственный интеллект	Глобальная машина Блока
14	Кибернадный тест	Может ли компьютер творить миры?	Дж. Баресси [Battesi, 1987]	Творение	Компьютер способен репродуцировать искусственных людей, общества, миры	Футуротехнологии

15	Тотальный тест	Может ли компьютер быть неотличимым от человека?	С. Харнад [Harad, 2001]	Человек и реальность	Компьютер способен дублировать человека	НБИКС
16	Самый тотальный тест	Может ли развитие компьютера не отличаться от естественной эволюции?	П. Швайзер [Schweizer, 1998]	Человек и общество	Компьютер способен к развитию, подобного социобиологической эволюции	НБИКС и природа
17	Тест личности	Может ли компьютер быть личностью?	Дж. Поллок [Pollok, 1995]	Личность	Компьютер может быть личностью	Артефакты проекта «Оскар»
18	Тест зомби	Может ли бессознательный компьютер казаться сознательным?	Р. Кирк [Kirk, 1971]	Бессознательное (внутри)/ сознательное (снаружи)	Философские зомби возможны	Компьютерные модификации художественных образов
		Могу ли я быть компьютерным зомби?	Д. Чалмерс [Chalmers, 1995]	Компьютерный зомби	Компьютерный Я-зомби возможен	Компьютер, соответствующий «новой теории информации»
19	Интроспективный тест	Может ли компьютер обладать «иным» сознанием?	А. Клифтон [Clifton, 2003]	«Иное» сознание	Компьютер способен обладать нечеловеческой психикой	Проект сенситивных компьютеров
20	Интерактивный тест	Может ли компьютер обрабатывать идеи?	В. Финн [Финн, 2009]	Идея	Рационализация общественной жизни посредством интеллектуальных технологий	Квазиаксиоматическая ДСМ-система
21	Комплексный тест	Может ли компьютер ВСЁ?	А. Алексеев [Алексеев, 2006]	Когниции частных тестов	Интеграция частных тестов Тьюринга	Комплексная машина Корсакова-Тьюринга

ГЛАВА 4. ДЕФИНИТНАЯ ФУНКЦИЯ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

Дефинитная функция оригинального теста заключалась в определении «интеллекта», который проявляется в условиях текстовой коммуникации и реализуется компьютерными средствами универсальной цифровой вычислительной машины и нейронной сетью. Дефинитная функция комплексного теста существенно расширяется. Теперь не только «интеллект» изучается в контексте компьютерной реализации. Исследуется возможность рационального определения когнитивных феноменов самого широкого спектра – витальных, ментальных, персональных, социальных. Когнитивные феномены рассматриваются с различных позиций. Это – мои личные мысли, приватные рассуждения, переживания, ощущения, квалиа и пр., т.е. *субъективные когниции*. Это – объективные теории естественных, гуманитарных, социальных, технических и смежных наук (*объективные когниции*). Это – атрибуции (приписывания) субъективных и объективных когниций способам и формам функционирования компьютерной системы (*когнитивно-компьютерные когниции*). Столь широкие обобщения примитивной схемы intersubjectивного взаимодействия людей и компьютеров, предложенной в оригинальном тесте Тьюринга в контексте экспериментальной обстановки, более того – концептуально воображаемой экспериментальной обстановки, требует раскрытия современного *реального социокультурного контекста*, в котором фактически, на самом деле происходят коммуникации подобного рода

Комплексный тест Тьюринга в контексте электронной культуры

В современном научном дискурсе используется ряд когнитивистских метафор культуры. «Киберкультура» ставит акцент на алгоритмически управляемых способах регуляции человеческого бытия. «Цифровая культура» спекулирует на универсальности дигитальной репрезентации мира. «Информационная культура» предлагает формат информационного процесса, системы, технологии. В «семиотической культуре» процесс становления, функционирования и трансляции феноменов культуры неотличим от динамики знаковых систем. «Компьютерная культура» предполагает метафору компьютерной системы, программно

манипулирующей «социокодами», которые вводятся, хранятся, обрабатываются и извлекаются из баз «данных», «знаний» и «смыслов» культуры. «Нейрокультура» переносит модели мозга на социум, метафорически отождествляя культурный процесс с нейрофизиологическим функционированием. «Медиакультура» исследует разнообразие средств и способов репрезентации.

С конца 90-х годов прошлого века в научном и общественном дискурсе часто употребляется термин «электронная культура» («э-культура», «e-culture»). Проводятся научные конференции, семинары, симпозиумы по проблематике электронной культуры (например, широко известная международная конференция «EVA»), при министерствах культуры различных стран создаются департаменты, отделы, центры (в нашей стране – «Центр ПИК»), в вузах организуются кафедры электронной культуры. Столь широкого применения не нашел ни один из приведённых выше когнитивистско-культурологических терминов. В свою очередь, трактовка «э-культуры» крайне обширна, полисемична. Поэтому одна из основных проблем э-культуры имеет лингвистическую окраску и характеризует трудность выработки унифицированного термина, концептуальное многообразие которого фиксируется чётким контекстом его употребления. То есть сегодня мы имеем немаловажную *лингвистическую проблему э-культуры*. Кратко суть решения этой проблемы в следующем. Выделяется ряд контекстов употребления термина, в которых «э-культура» приобретает определённое значение. Далее исследователь произвольно выбирает требуемый контекст из нижеперечисленных.

1. *Технологический контекст*. Первая дефиниция э-культуры не имеет референта и в высшей степени абстрактна: э-культура – всё то, что посредством электронных технологий способствует сохранению и умножению материальных и духовных благ человека, общества, человечества. Вторая дефиниция более конкретна. Предполагается наличие некоторой априорной культурологической концепции и возможность её реализации электронными технологиями: э-культура – это культура, основанная на достижениях интегративного комплекса НБИКС-технологий (NBICS: нано-, био-, инфо-, когни-, социокультурной технологии). Правомочность дефиниции обусловлена тем, что комплекс НБИКС имеет единый компьютеринговый базис редукции к регулярным методам конструирования электронных комплексов (нано- и био-

технологии), обработки электронных носителей информации (инфо- и когнитотехнология) и конструктивного использования результатов электронных преобразований (социокультурная технология).

2. *Эпистемологический контекст.* Электронная культура – это культура, в которой духовная и материальная составляющие формируются технологией искусственного интеллекта – технологией извлечения, репрезентации, хранения, обработки, передачи «знаний». Справедливость данного определения обусловлена тем, что основа любой технологии – интеллектуальная деятельность. Аналогично этому, электронная «компонента» интеллектуальной деятельности – речь идет об э-культуре – искусственный интеллект. Следует более чётко определить понятие ИИ, так как в условиях э-культуры существенно конкретизируются исходные идеи полувековой давности по поводу компьютерной реализации (имитации, моделирования, воспроизведения) абстрактных свойств и способностей типа мышления, памяти, интуиции, сознания, творчества, обучения и пр. Современные исследования ИИ сфокусированы на проблемно-ориентированных интегративных областях. С начала 21 в. институционально закрепляются следующие направления ИИ: искусственная жизнь, искусственная личность, искусственное общество, искусственный мир. Искусственная жизнь тесно связана с исследованиями биотехнологии и ориентирована на компьютерную реализацию феномена биологической жизни. В когнитологических проектах искусственной личности акцент делается на имитации, моделировании или воспроизведении морального поведения систем, которые внешне оцениваются как обладающие свободной волей самосознающие индивиды. В исследовании искусственных обществ – компьютерного фундамента информационных и социокультурных технологий – воспроизводятся феномены общественной жизни, начиная от материально-производственной сферы и завершая религиозными верованиями. Направление «искусственный мир» сегодня пока недостаточно развито, однако заявлен ряд претензий, осуществление которых соотносится с завершением процесса интеграции комплекса НБИКС. Всем данным исследованиям не только согласно исторической памяти присваивается ярлык «искусственный интеллект». В основе проектов – творческая деятельность человека, естественные интеллектуальные способности которой поддерживаются и усиливаются инструментарием электронных технологий.

3. *Репрезентационный контекст.* Задается современными возможностями репрезентации феноменов традиционной культуры. К. Вельтман, один из пионеров определения «электронной культуры», связывает э-культуру с традиционными объектами культуры и искусства, репрезентированные средствами информационно-коммуникационных технологий (ИКТ). Сюда включаются электронные библиотеки, виртуальные музеи, мультимедийные реконструкции исторических памятников, «семантическая сеть» в среде интернет, систематизирующая факты культурологии, археологии, этнографии, искусствознания, историографии и пр.

4. *Виртуалистский контекст.* Э-культура – качественно новая реальность, не имеющая натуральных аналогов. Приложения оцифрованной информации не сводятся к инструментальному использованию технологий. Компьютеры, мультимедиа и интернет открывают новые формы выражения человека, изменяют роли и статусы институтов культуры и искусств. Например, по мнению А. Рончи, виртуальный музей – это не просто коллекция графических образов культурного наследия, воспроизведённых на компьютере. Это – новая форма интерпретации исторической реальности, которую «проживает» посетитель музея.

5. *Социокультурный контекст.* Э-культура не сводится к применению и распространению новых ИКТ. Это – не техногенное явление, а социокультурный феномен, влияющий на ментальность, культуру в целом, способы восприятия, трансляции социокультурной информации. Прежде всего – это появление новых и коренная модификация существующих социальных отношений, культурных ценностей, норм, стереотипов поведения, в которых начинают доминировать компьютерно-технологические формы деятельности. Э-культура – часть современной культуры с присущими информационному обществу специфическими мировоззренческими и нравственными универсалиями (К.К. Колин).

6. *Антропологический контекст.* Э-культура – это культура человека нового вида, е-Номо. Е-хомо сохраняет биологическую принадлежность к Homo Sapiens, но качественно отличается от него за счет симбиоза с продуктами стремительно развивающихся сверхвысоких технологий НБИКС. Среда «э-обитания», естественный «э-отбор», «роботы внутри нас», «э-свобода» и пр. – это параметры э-культуры э-

человека. Э-человек появится при жизни основной части нынешнего поколения (А.С. Нариньяни).

7. *Институциональный контекст.* Социальная кластеризация э-культуры отличается от структуры традиционной культуры, хотя прослеживаются функциональные аналогии. Э-культура – это совокупность социальных институтов, организуемых средствами технологий: э-образование (e-learning), э-политика (e-polity), э-наука (e-science), э-торговля (e-commerce), э-медицина и др. (М. Шварц). Приставка «э-» применяется к тем сферам общественной жизни, в которых при необходимости может быть задействован инструментарий компьютерной обработки сложных информационных процессов и поддержки разветвленной сети коммуникации. Институциональный контекст необходим для изучения феномена э-культуры, так как непосредственно внедрён в практику. Мы непосредственно соприкасаемся с феноменами э-культуры в нашей повседневной жизни, поэтому есть на что опереться в изучении феномена э-культуры. Проблемы э-культуры очевидны с позиции здравого смысла, они касаются существенных сфер человеческого бытия. Предыдущие контексты употребления термина представляют несомненный интерес, но в наибольшей мере они обладают теоретической значимостью для проведения научных исследований.

Дефиницию когнитивных феноменов в контексте э-культуры будем изучать, отталкиваясь от схемы интересубъективного взаимодействия людей и компьютеров, которая была приведена во введении. Вначале исследованию подлежит дефиниция социального феномена, проявляющегося при нашем общении с некоторым человеком или коллективом людей, которые находятся по ту сторону стены Тьюринга и которые, возможно, людьми-то и не являются. Далее вопрос специфицируется – в ходе тьюрингового тестирования *Я* должен идентифицировать *другого* как личность. Это значит, что основным условием тьюринговой дефиниции когнитивного феномена выступает «сознание» как главный атрибут личности. Рассмотрим, каким образом комплексный тест Тьюринга способствуют проведению данной работы, формируя в компьютерном контексте дефиниции социальных, персональных и ментальных феноменов.

4.1. Дефиниция социальных феноменов: «другой» в электронной культуре

Проблема *другого* как социально-философская проблема

В современной философии обозначим несколько подходов к *проблеме интерсубъективности*, которая ранее называлась «*проблема другого*» или, в зависимости от конкретики исследовательского интереса, – «*проблема другого сознания*»¹: 1) социологический подход ставит вопросы о включённости мира других в жизненный мир *Я* и взаимовлиянии этих миров; 2) эпистемологический подход направлен на обоснование наших суждений, выражающих убеждённость о другом и саму возможность наличия другого; 3) технологический подход (предлагается автором) ставит вопросы о возможности сознания компьютерной системы, способной имитировать, моделировать, репродуцировать, креацировать *другого* в условиях прохождения этой системой комплексного теста Тьюринга.

Социологическая проблема *другого*

Социально-философская проблематика *другого* обязана в первую очередь феноменологии оснований осведомлённости о *другом*. Проблема связана с человеческой субъективностью, человеческой идентичностью, нашим опытом себя, нашим опытом других, нашими отношениями к другим, а также с тем, как другие воздействуют на нас и как другие существенно включаются в наше ощущение себя.

Э. Гуссерль признаётся основателем феноменологической проблематики *другого*. В рамках трансцендентальной философии он делает набросок интенционального истолкования *другого* (см. «Картезианские размышления», «Размышление пятое»). Утверждается, что *Я* – опыт объективного мира – имеет в то же время опыт *других*, так как *другой* человек сам «*живе*» стоит передо мной. Восприятие *другого* возможно

¹ В основе данного исследования лежат работы Д.И. Дубровского, посвященные проблеме «*другого сознания*» и концентрированно представленные в [Дубровский, 2006]. Автор пользовался этими положениями в ряде работ, например [Алексеев, 2008-3; 5]. Проблема интерсубъективности терминологически не будет отличаться от «*проблемы другого*». Второе наименование проблемы – «*проблема другого сознания*» имеет внушительные корни и традиции, так как обсуждается начиная, как минимум, с работ Декарта (см. выше тест Декарта).

в результате аппрециации (приведения-в-со-присутствие) и уподобляющей апперцепции, которая не есть вывод по аналогии, а интенциональное схватывание тела *другого*, которое обладает модусом «там», в отличие от моего психофизического единства, обладающего модусом «здесь». Продуктивно для нашей темы суждение об аномальной апперцепции – о восприятии глухих, немых, высших животных – относительно нормального случая, т.е. относительно *Я*. В таком случае аномальной модификацией моего человеческого существования выступает не только глухой, немой, косой человек, но и компьютер, проходящий тест Тьюринга.

М. Хайдеггер исходит из онтологической перспективы, в соответствии с которой *другие* представлены в форме существования каждого из нас. Они являются в центре нашего способа существования. Мы пребываем в мире, и этот мир является миром, конституированным другими. Наше бытие является бытием с другими в большей мере, нежели со знанием о других.

Ж.-П. Сартр принимает точку зрения (см. «Бытие и ничто», ч. 3, гл. 1 «Существование другого»), согласно которой другие люди востребованы нами в приобретении нашего собственного смысла себя как личностей. Беспокойство «для-меня» открывает мне бытие, которое есть мое бытие, не являясь-для-меня. Наш смысл *других* кажется предшествующим нашему смыслу себя. Если и имеется асимметрия между *Я* и *другим*, то это – реверсивная, обратная асимметрия. Мы приходим через *другого* в нашем собственном опыте и в этом опыте критически смотрим на него. Наше бытие имеет *других* в своём центре. Существование *другого* переживается с очевидностью в факте и через факт моей объективности. Этот объект, каким *другой* является для меня, и объект, которым *Я* являюсь для *другого*, обнаруживаются как тела. Однако для нас не тело является первым и не оно нам раскрывает вещи, а вещи-орудия в своем первоначальном появлении указывают на наше тело. Здесь также возможны интуиции относительно вещей-орудий, каковыми являются и компьютерные системы.

М. Мерло-Понти посвящает проблеме *других* сознаний материал «Феноменологии восприятия» своей работы «Другие и человеческий мир». Он утверждает, что каждый из нас ощущает мир как общий с *другими*. Наш опыт мира зависим от опыта социального мира. Фундамен-

тальным обращением к другим сознаниям является психофизическая целостность. Восприятие человеческого тела в поступке – это непосредственное восприятие личности. Однако *Я* не более определён в моём собственном существовании, нежели чем *Я* определён в существовании *другого* (*других*).

П. Рикёр, осуществляя «герменевтическую прививку» к феноменологии, что влечёт не только переосмысление теории лингвистических значений, но в большей мере проблематику *Cogito*, полагает, что интерпретатор социокультурных явлений как текстов может присвоить себе смысл *другого* (автора текста), из чужого делаая его своим. Благодаря этому расширение самопонимания достигается через понимание *другого*, более того, всякая герменевтика выступает пониманием самого себя через понимание *другого*.

В социологическом подходе фиксируется, что мы находимся в зависимости от *других* не просто с позиции нашего существования, но в большей мере для осмысления самого себя, наша осознанность *других* является сердцевинной нашей осознанности себя. Имеется и диаметрально противоположная позиция – мы никогда не можем осознать опыт другого существа. Наша самость огораживает нас от *других*, мы можем знать лишь собственную внутреннюю жизнь, но не жизнь *других*. Между нами и *другими* – фундаментальная пропасть.

Несомненно, феноменологические исследования заслуживают пристального внимания в связи с нашей темой. Важную большую работу в этом направлении, как раз в контексте изучения теста Тьюринга, недавно проделала Н.М. Смирнова [Смирнова, 2005-1; 2; 2006; 2011]. В частности, она полагает, что конститутивным принципом *другого* в условиях компьютерных коммуникаций является не аналогия кинестетических движений, но аналогия наделяния смыслом [Смирнова, 2005-1]. Следует добавить, что такого рода «смысл» в условиях компьютерных технологий следует репрезентировать мультимедийными средствами [Алексеев, 2004].

Эпистемологическая проблема «другого»

А. Гуслоп считает, что эпистемологическая проблема возникает в силу радикального различия между нашим доступом к собственному опыту и доступом к опыту других людей [Hyslop, 2005; 1976; 1998]. Мы не всегда непосредственно знаем, что пребываем в ментальном состоя-

нии, в котором пребываем, но что совершенно очевидно, мы никогда не имеем прямого знания о том, что другие люди пребывают в каком-нибудь ментальном состоянии, в котором они пребывают. Имеется *радикальная асимметрия* – не в том, что можно наблюдать, воспринимать, чувствовать как противоположное тому, что невозможно наблюдать, воспринимать, чувствовать. *Я* могу понять ментальные состояния другого человека, но это не значит, что проблема других сознаний не стоит передо мной. Что мне нужно – это способность переживать ментальные состояния другого человека как свои собственные состояния. Со мной должно случаться то, что случается с *другим*.

Эпистемологическая проблема разрабатывается в русле аналитической философии и в большей мере связывается с проблемой *других сознаний*. Проблема стала считаться философской с 20-го века, когда Джон Стюарт Милль предложил *умозаключение по аналогии к другим сознаниям*. Суть умозаключения в следующем: основным предметом исследования должна быть каузальная связь между ментальными состояниями и нашим поведением. Умозаключение по аналогии к другим сознаниям удерживается до середины 20-го столетия. Это умозаключение предназначено для обоснования достоверности знаний о внутренней жизни других людей по аналогии с моей внутренней жизнью. Другие люди очень схожи со мной. Их поведение очень схоже с моим поведением, осуществляемым мною в подобных условиях. *Другие* «сделаны» из той же материи, что и я. Когда *Я* обжигаюсь, мне больно, я вскрикиваю и вздрагиваю. Когда *другие* обжигаются, они делают то же самое. Отсюда *Я* могу заключить, что они также испытывают боль. Имеется много таких уподоблений. В общем случае, *Я* знаю о том, что обладаю убеждениями, эмоциями, ощущениями, чувствами и соответствующим им поведением. Поэтому, основываясь на многообразных основаниях аналогии в поведении, *Я* способен заключить, что *другие* также, как и я, имеют убеждения, эмоции, ощущения и пр. Короче говоря, *Я* получаю право на умозаключение о том, что *другие* обладают тем же, чем и *Я* обладаю, т.е. внутренней жизнью, и что их жизнь очень схожа с моей жизнью.

Однако умозаключение по аналогии неадекватно, так как, во-первых, оно не только неверифицируемо, но и недедудируемо, т.е. логический вывод невозможно ни обосновать, ни продемонстрировать. Во-вторых, аргументы для индуктивного обобщения базируются только

на единственном случае – т.е. на уникальном случае наших ментальных состояний и поведения. Данное умозаключение принять трудно. Сегодня более распространена гибридная форма умозаключения, включающая научное умозаключение и гипотетическое умозаключение, однако проблемы с обоснованием остаются прежними.

А. Гуслоп обсуждает также иные подходы. Некоторые отрицают, что имеется проблема, утверждая, что ментальные состояния других вовсе не скрыты от нас. У нас есть прямая осведомлённость о таких ментальных состояниях. Поэтому мы имеет прямые знания и проблема отсутствует. Конечно, это неправдоподобно. Имеются также послабления радикальной асимметрии. Так, представители феминизма считают, что асимметрия на самом деле присуща мускулинным индивидам. Именно мужчины полагают, что имеется проблема с нашим опытом и опытом *другого*. Женщины, как правило, чувствуют себя в согласии с *другими*, для них эта проблема не возникает.

Имеется ещё один подход – проблемы *другого* вообще нет. На это А. Гуслоп отвечает, что нет такой теории сознания, которая может избежать проблемы других сознаний.

Любая теория сознания включает теорию других сознаний
--

Технологическая проблема *другого*

При изучении технологической проблемы *другого* имеются в виду философские аспекты когнитивно-компьютерной технологии. К ним относятся позиции различных теорий сознания относительно воспроизводства феномена сознания в компьютерной среде. Они вырабатываются в физикализме, бихевиоризме, функционализме, элиминативном материализме, эпифеноменализме и др. Подчеркнем, что эти теории сознания выступают методологической базой для суждений о возможности прохождения *x*-системой теста Тьюринга.

Бихевиоризм. В решении проблемы *другого* нет трудностей. Основополагающей является мысль о том, что бытие в ментальном состоянии является способом, которым некто ведёт себя или обладает диспозицией к поведению. Отсюда нет большой трудности *знать* о поведении и даже поведенческой диспозиции *другого*. А это знание можно операционализировать средствами компьютерной техники. Ранее бихевиоральная позиция затрагивалась при анализе теста Блока.

Функционализм. Проблема других сознаний имеется, но она представляет собой лишь небольшое затруднение. Различные ментальные состояния характеризуются различными ролями, посредством которых организм отвечает на своё окружение, их типичными причинами и следствиями. Поэтому боль от ожога является таким внутренним состоянием, которое обычно причиняется ожогом и, как правило, приводит к содроганию, плачу и тому подобному поведению. Отсюда следует, что другие человеческие существа имеют схожие внутренние состояния, просто надо внимательно наблюдать, как они ведут себя и в каком окружении. Сознание является вычислительным устройством и на теоретическом уровне отождествляется с реализацией в машине Тьюринга. Вычислительное устройство отождествляется с мозгом, на котором «запущена» программа. Несомненно, функционализм – это наиболее плодотворная теория сознания с точки зрения компьютерного моделирования или, по крайней мере, компьютерной имитации *другого*.

Элиминативный материализм. В концептуальных структурах здравого смысла о ментальных понятиях воплощено ошибочное представление о соответствующих феноменах сознания. Каких бы успехов ни достигла когнитивная наука, понятия здравого смысла нельзя будет редуцировать или прояснить с помощью более совершенной теории. Поэтому эти понятия следует элиминировать и создавать термины когнитивно-компьютерной науки. Здесь прослеживается реверсивная асимметрия – мы начинаем объяснять то, что нам кажется болью, желанием, надеждой, мышлением и пр. в операционально-функциональных терминах компьютерной системы. Само же понятие сознания, как оно сейчас трактуется, вытесняется новой терминологией. Короче, любовь – это поток Q-данных из E-регистра в F-регрстр.

Физикализм. Сопротивляется идее чисто компьютерного воспроизводства феномена сознания. Требуется физическая основа, тождественная с основой оригинала. Здесь скорее проблема *другого* возникает в условиях биологического клонирования. Физическая субстанция идентична. Только является ли полученное в результате опыта существо искомым «другим», наделённым всем спектром биологических, персонологических, социальных свойства и способностей? С другой стороны, генная инженерия – это очень сложная компьютинговая система.

Эпифеноменализм. Сознание – это призрачный «довесок» к физическим процессам. Ментальные события причиняются физическими событиями в мозге, но не оказывают никаких эффектов на какие-либо физические события. Поведение причиняется мышцами, которые связаны с поступающими нейральными импульсами, а нейральные импульсы генерируются входом от других нейронов или от органов чувств. Ментальные события не играют каузальной роли в этих процессах. Проблема с воспроизводством сознания на компьютерной основе исключается – ведь и моё сознание эффекторно не выразимо, не говоря уже о «другом» сознании.

Комплексный тест Тьюринга как дефиниенс *другого* в компьютерных коммуникациях

Поиск теоретических оснований для убеждённости в том, что вы общаетесь с человеком, а не с компьютером, упирается в тест Тьюринга – если тестируемая система не проходит тест Тьюринга и его модификации, то перед вами – не человек. Однако очевидно, что не все люди пройдут тест Тьюринга на творческие способности, на понимание, на субкогнитивные особенности и пр. Что уж говорить по поводу игры в шахматы – ведь не у каждого способности гроссмейстера. Более того, выделение компьютерного и человеческого в естественной системе – это искусственный приём, в реальности человек представляет собой единство механического – того, что представимо вычислительными схемами, и немеханического – того, что не поддаётся компьютеризации. Любая система искусственного интеллекта – это репрезентация, выражение естественного интеллекта разработчиков, экспертов, программистов. Если учесть социальный контекст разработки, т.е. обусловленность знаний разработчиков от общественных форм сознания, то можно утверждать, что в создании компьютерной системы участвуют культурные эпохи, территориальные континенты и поколения разработчиков.

В данной работе приведены различные позиции относительно проблемы *другого* в контексте э-культуры. Следует отметить две противоположные точки зрения относительно того, что же получается в итоге когнитивно-компьютерных новаций. Первая позиция достаточно прозрачно выражается Д.И. Дубровским в целом ряде работ [Дубровский, 2006; 2007; 2011]. Суть позиции в том, что, создавая компьютерную систему, функционально эквивалентную организации сознательного

существа, в перспективе на выходе получаем сознательное существо. В философии искусственного интеллекта такую компьютерную систему принято называть «искусственной личностью». Моя точка зрения противоположна – мы создаем зомби, т.е. квазисуществ, которые функционально, поведенчески и даже (в условиях нанотехнологий) физически подобны нам, людям, но при этом сознанием не обладают. Это – не истинные философские зомби, у них имеется блок «псевдосознания» (см. ниже). Таким образом, проблема *другого* в контексте э-культуры упирается в проблему «зомби»: с кем *Я* общаюсь по ту сторону чата, с человеческой личностью или с её функциональным симулякром?

Технологическая проблема *другого* в отечественных междисциплинарных исследованиях ИИ соотносится с общей социально-философской проблемой интерсубъективности, содержанием которой является «экспликация теоретических оснований репрезентации социального – «логоса социальности» как интерсубъективности высшего порядка... Это изучение взаимной интенциональности участников коммуникации... и полагает обусловленность продукта когнитивной деятельности его инструментальными характеристиками» [Смирнова Н.М., 2011, С. 110].

В контексте исследований ИИ «условия возможности социальных коммуникаций» [Смирнова Н.М., там же] соотносятся с компьютерным инструментарием и определяются его способностью имитировать и/или репродуцировать первозначимые параметры *другого* – его личностные, персонологические характеристики. Рассмотрим, насколько успешно справляется с «личностью» современная методология искусственного интеллекта.

4.2. Дефиниция персональных феноменов: «личность» в электронной культуре

Проблема дефиниции «личности» – самосознательного активного морально вменяемого субъекта, обладающего всем спектром качеств и способностей, которые культура закрепляет за термином «личность», крайне сложна в контексте традиционной культуры. Ещё большую сложность вызывает изучение «личности» в электронной культуре, в условиях информационно-компьютерных коммуникаций. В новом контексте телесная идентификация может напрочь отсутствовать либо подменяться симулякрами виртуальной реальности. А телесность, как

мы видели выше, для социальных философов традиционной культуры является главным атрибутом *другого*, «личности». Если тестировать все возможные параметры «личности» в условиях комплексного теста Тьюринга, т.е. проводить положительную дефиницию тестируемых персонологических качеств и способностей, то получается слишком обширный базис для индуктивного заключения – да, *Я* общаюсь с личностью, либо, напротив, со мной коммуницирует веб-робот. Поэтому предпримем стандартный ход, который применяется в доказательстве от противного (*reductio ad absurdum*).

Требуется доказать суждение «*Я общаюсь с личностью*» (*p*). В начале доказательства формулируется противоположное ему высказывание не-*p* и предполагается, что оно истинно. Пропозициональное суждение не-*p* означает по сути то, что *Я* общаюсь с не-личностью, т.е. неизвестная система *x* находится по ту сторону стены Тьюринга – это некий философский зомби. Если допускается, что *p* ложно, тогда должно быть истинным не-*p*. Затем из этого якобы истинного антитезиса следует выводит следствия – некоторые суждения $p_1, p_2, \dots, p_n$ – до тех пор, пока либо не получится противоречие, либо такое следствие, которое явным образом не соответствует очевидным фактам. То есть мы получаем, что если некоторое $p_1, p_2, \dots, p_n$ ложно, то не-*p* истинно.

Например, если воспользоваться тестом Блока (нео-ТТ), то в предположении общения с зомби – а именно так я должен начинать свое исследование партнера в социальной сети – я должен получать бессмысленный набор символов, хотя эти наборы могут быть синтаксически корректными, а в семантическом отношении что-то отражать. Сообщение p_1 – бессмысленно. Сообщение p_2 – бессмысленно и пр. Однако вдруг поступило по интернету высокохудожественное сообщение p_n , проникнутое глубоким смыслом и морально-духовным содержанием. Оно очень даже осмысленно, поэтому логично предположить, что *Я* общаюсь не с зомби, а с «личностью».

Если же теперь воспользоваться прямым способом аргументации, то цепь моих рассуждений должна быть такой: *Я*-судья общаюсь с «личностью», поэтому мне должны поступать по интернету высокохудожественные сообщения, проникнутые глубоким смыслом и морально-духовным содержанием. Мне поступают сообщения: p_1 – высокохудожественное, p_2 – глубоко-нравственное и пр. Все эти сообщения я

должен анализировать на предмет «высокого» содержания. Одно из поступивших сообщений – бессмысленное, поэтому я общаюсь с зомби. Однако этот единичный факт статистически сомнителен и требует дополнительного пополнения базиса индукции.

**Методология философских зомби экономнее стандартного
метода индуктивного подтверждения личностных параметров
другого**

Рассмотрим проблему зомби подробнее и выясним, каким образом она способствует решению проблемы «личности» в электронной культуре. Тема философских зомби представляется родной для автора, так как он полагает, что впервые концентрированно «породил» её в контексте отечественной науки [Алексеев, 2005–2008].

Следует отметить, что в аналитической философии в течение последних десятилетий возникло интригующее направление исследований, названное «проблемой зомби». Под философскими зомби в общем случае подразумеваются бессознательные системы, которые поведенчески, функционально и/или физически тождественны, неотличимы и/или подобны сознательным существам. Проблема философских зомби обширна, многогранна, многопланова. За последние тридцать лет вышли десятки монографий и сотни крупных статей именитых зарубежных авторов. Даже сложилась классификация исследователей. Например, *зомбифилами* называют тех, кто принимает тематику зомби для критики или обоснования теорий сознания. *Зомбифобы*, напротив, тематику зомби игнорируют.

Аргумент философских зомби

В проблематике зомби наиболее важен *аргумент зомби*. В общей форме он формулируется условно-категорическим заключением (*modus ponens*): 1) если зомби возможны, то некоторая теория сознания ложна; 2) зомби возможны; 3) следовательно, данная теория сознания ложна.

Возможность зомби выступает в разных формах. Рассматриваются логическая, онтологическая, метафизическая, номическая, физическая, эмпирическая, риторическая и прочие разновидности возможности. Как показала практика философских дискуссий, этих модальностей недостаточно. Поэтому сегодня в проблематике зомби более фундаментальным считается *аргумент мыслимости зомби*. Он представляется в

форме силлогизма: 1) *зомби мыслимы*; 2) *всё мыслимое возможно*; 3) *следовательно, зомби возможны* [Kirk, 2006].

Но и этого оказалось недостаточно. «Мыслимость» понимается в дифференцированной форме. Так, Д. Чалмерс применяет идеи крипкеанской «двумерной семантики» к понятию мыслимости, выделяя априорную и апостериорную мыслимость зомби [Chalmers, 1993-1; 2]. Его последователи отмечают ещё ряд градаций – «*n*-мыслимость». Изучение модальных аспектов проблемы зомби особая тема. Здесь же хочется обратить внимание на разделение лагеря зомбифилов. Будем различать стратегии зомбистов, антизомбистов и нейтральных зомбистов.

Позиция зомбистов, антизомбистов, нейтральных зомбистов

Зомбисты утверждают возможность зомби или, по крайней мере, их мыслимость. Если, например, возможны поведенческие зомби, то ложен бихевиоризм; если возможны физические зомби, то ложен физикализм. Имеются и другие фундаментальные выводы, например: 1) эпифеноменализм истинен; 2) проблема *другого* неразрешима, так как мы оцениваем не субъективную реальность *другого*, а лишь внешнее поведение, т.е. оцениваем его двойника-зомби; 3) сознание не существенно для эволюции природы; 4) изучение сознания для прогресса когнитивно-компьютерных технологий не интересно.

Антизомбисты заявляют противоположное: идея зомби настолько абсурдна, что зомби не только вообразить нельзя, но и более-менее связано представить. Поэтому по причине немыслимости зомби аргументы зомби не опровергает ту или иную теорию сознания, например материализм, а напротив, защищают эту теорию. Позиция четко представлена в [Marton, 1998] – идея зомби критикуется с позиции физикализма.

Нейтральные зомбисты требуют внести ясность в проблематику сознания, распутать сложный «клубок из терминов», большинство которых не определимо, полисемично, нередуктивно и чётко определяется лишь в оппозиции к контрадикторному понятию, т.е. к понятию бес-сознательного. Над теоретическими проблемами метафизики сознания не стоит задумываться, а нужно обратить внимание на роль сознания в жизни человека и в эволюции природы, практически оценивать когнитивно-компьютерные механизмы воспроизводства квазисознательных феноменов на субстрате, отличном от биологического, и пр.

Отечественному философскому сообществу пока не грозит подобного рода зомбиевый сепаратизм. Для нашей философии проблема зомби вне связи с философией ИИ вообще не стояла. Впервые тематика философских зомби как специфическая тема философских исследований отечественной науки прозвучала в докладе автора на конференции «Философия искусственного интеллекта» (Москва, 18 – 21 января 2005 г.) [Алексеев, 2005-3], в докладе на IV Российском философском конгрессе (Москва, 24 – 28 мая 2005 г.) [Алексеев, 2005-11] и в последующем была представлена в ряде его статей [Алексеев, 2008-2; 2009-3]. Перед этим высказывались отдельные суждения по этой теме в некоторых работах Д.И. Дубровского [Дубровский, 2004]. Также был проведён детальный анализ проблематики зомби у Д. Деннетта в связи с популяризацией его работ Н.С. Юлиной [Юлина, 2004].

В 2007 г. тема зомби окончательно «прорвалась» в отечественную философию – в «Философских науках» (№ 3, 2007 г.) вышла статья С.Ф. Нагумановой, названная «Аргумент мыслимости против материализма», в которой обсуждается обозначенный выше «аргумент мыслимости зомби», а не мыслимости «вообще» [Нагуманова, 2007]. Следует дополнить данную интересную статью: цель состоит в попытке систематического *определения философских зомби*, иначе вышеуказанная статья, в силу неопределённости зомби в отечественной литературе, становится непонятной и аргумент мыслимости зомби определится как «х через у».

Становление теста Зомби

Понятие «философский зомби» исследователи начинают определять контрастным способом, выделяя его среди несравнимых терминов: «Зомби» – это тупой человек; чудной тип; новобранец; коктейль из рома с содовой; пост-панк группа; UNIX-процесс, «вхолостую» использующий вычислительные ресурсы; серия компьютерных игр и пр. [Chalmers, 2013; Polger, 2003]. К этому списку несколько лет назад добавилась «компьютерный зомби» – разновидность диверсионных программных средств, актуализирующихся по указанию хакера для засорения интернета спамом.

Образы философских зомби инспирированы в большей мере голливудскими и гаитянскими зомби. *Голливудский зомби* – это что-то вроде реанимированного трупа из фильмов ужасов (B-movies), который по отношению к человеку ведёт себя агрессивно, поедая его плоть. Данно-

му образу более семидесяти лет, начиная с проката первого фильма («Белые зомби», 1933 г.). Ежегодно такой образ обрастает нюансами из-за появления десятков новых зомбиевых фильмов-ужасов. *Гаитянский зомби* – существо, управляемое чужой волей и/или лишённое души за счёт зомбификации – процедуры заклинания или принятия снадобья. Образ голливудского и гаитянского зомби чётко прослеживается соответственно в первом и втором мысленных экспериментах Р. Кирка (см. ниже).

Несомненно, через обиденное сознание образы столь эксцентричных мифических «существ» проникли в воображение и интуицию зомбифилов. Так, Л. Хаусер ярко живописует: «Если голливудские зомби поедают мясо, [то философские] зомби питаются дуалистической материей: на завтрак у них – квалиа. К несчастью для функционалистов, на ленч у них – программы. А к прискорбию для теории тождества, на обед у них – мозги» [Hauser, 1995].

Следует отметить, что слово «философские зомби» имеет ещё одно курьёзное значение, а именно, это – зомби, обсуждающее философские темы [Balog, 1999].

Под влиянием столь выразительных образов зомбифилы определяют понятие «философского зомби». Некоторые, однако, спрашивают «Можно ли вообще зомби помыслить положительно?» и, отрицательно отвечая, полагают, что понятие «зомби» в общем-то ни к чему и не применимо [Masrou, 1998]. Явные определения дают многие другие исследователи: И. Арануси, Э. Бейл, Д. Дубровский, Н. Кертик Р. Кирк, А. Ленайер, Д. Ллойд, Ф.С. Масроу, П. Мертон, Т.С. Моуди, Т. Нагель, Т. Полджер, О. Фланеган, С. Харнад, Л. Хаусер, Д. Чалмерс и др.¹ В

¹ Kirk R. (2005) *Zombies and Consciousness*. University of Nottingham, November, 2005. (<http://www.oup.co.uk/isbn/0-19-928548-9>); Kirk, R. (2004). *Zombies*; Дубровский Д.И. (2003) Новая реальность: человек и компьютер / Полигнозис № 3 (23), С. 20 – 32; Дубровский Д.И. Проблема духа и тела: возможности решения (в связи со статьей Т.Нагеля «Мыслимость невозможного и проблема духа и тела»). Вопросы философии, 2003 г.; Nagel, T. (1998). *Conceiving the impossible and the mind-body problem*. *Philosophy* 73:337 – 52; <http://www.nyu.edu/gsas/dept/philo/faculty/nagel/papers/conceiving.pdf>; Chalmers, D. J. (2004). *Zombies on the web*; Hauser Larry (2001), *Zombies Invade Philosophy!*; <http://members.aol.com/lshauser/zomboid.html>; Polger Tom (2003), *Zombies*; Lanier Jaron (1995), "You can't argue with a zombie";

целом, вводимые этими исследователями дефиниции зомби совпадают с нашим общим определением. Имеется, конечно, ряд дополнений. Например, для Р. Кирка важно подчеркнуть в сознании точку зрения «каково быть». «Ничем не быть» или, точнее, «каково это – ничем не быть» – это «быть зомби». Д. Дубровский считает зомби сложным самоорганизующимся устройством, который обладает всеми функциональными свойствами человеческого организма, но начисто лишён внутренней, субъективной реальности. Т. Нагель подчёркивает мыслимость зомби, но невозможность в строгом смысле слова «возможности», и противопоставляет зомби «искусственной личности» (роботу, обладающему квазисознанием). По поводу мыслимости зомби с ним солидарен Ф.С. Масроу, считающий, что из знаменитого аргумента зомби следует метафизическая возможность миров зомби, однако зомби можно мыслить лишь негативно. Также с Т. Нагелем солидарен С.Харнад, но уже по поводу оппозиции «зомби/искусственная личность», полагая, что робот, функционально неотличимый от нас, людей, не может быть бессознательным зомби. Д.Чалмерс указывает на эвристическую роль зомби, считая их гипотезами, которые творчески вдохновляют философов. Он использует в качестве заменителя дефиниции смутную метафору – у зомби «внутри всё темно». Более чётко позицию самого же Д. Чалмерса описывает его ученик И. Арануси: зомби – это физический дубликат меня, поэтому он должен быть и моим функциональным дубликатом.

<http://www.well.com/user/jaron/zombie.html>; Bailey Andrew. Phisicalism and the Preposterousness of Zombies (Running head: The Preposterousness of Zombies). Department of Philosophy The University of Guelph Guelph; Aranyosi István A. (2004). Chalmers's Zombie Argument; Lloyd Dan (1997), Twilight of the zombies. (<http://www.trincoll.edu/~dlloyd/twilight.htm>); Kartik Navin. In The Hands of Zombies. (<http://www.stanford.edu/group/dualist/vol7/pdfs/kartik.pdf>); Masrou Fari S. Are Zombies Positively Conceivable?; Flanagan Owen, Polger Tom (1995), "Zombies and the function of consciousness", Journal of Consciousness Studies 2:313-21; <http://homepages.uc.edu/~polgertw/Polger-Zombies JCS.pdf>. См. перевод Т.А. Кураевой в кн. Тест Тьюринга. Роботы. Зомби: Пер. с англ. / Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с.; Marton, P. (1998). Zombies vs. materialists: The battle over conceivability. Southwest Philosophy Review 14:131-38; <http://www.brown.edu/Departments/Philosophy/zombie.html>; Moody Todd (1995), "Conversations with zombies"; http://www.imprint.co.uk/Moody_zombies.html

Л. Хаусер указывает на деструктивную функцию зомби, так как они разрушают проработанную материалистическую философию сознания и научную психологию.

А. Ленайер считает зомби дополнительным приёмом, «прикормкой» в дискуссиях по поводу проблемы «психика/тело» и исследования сознания. К нему примыкает Э. Бейл, подчёркивая сугубо теоретический и по сути технический характер этого понятия в дискуссиях по проблеме сознания.

О. Фланаган и Т. Полджер называют зомби «несчастливым болваном», который воюет то на одной, то на другой стороне в философских баталиях по сознанию. Однако они указывают на продуктивность этой темы, так как проблема зомби предельно заостряет вопрос о роли сознания, раскрывает несостоятельность функционализма, опровергает тест Тьюринга и демонстрирует неразрешимость традиционной проблемы «других сознаний» – как можно быть уверенным в том, что некоторые, а может быть, и все окружающие нас люди – не зомби? Т.С. Моуди, считая зомби функционально полной и детальной дескрипцией познавательной деятельности, т.е. бесчувственным симулякрот сознательного существа, полагает, что проблема зомби – это весьма полезная вариация на тему «других сознаний» и яркая концептуализация философских вопросов о сознании. Д. Ллойд «проблему другого» связывает с критерием «зомбиевости» и полагает, что этот критерий влечёт за собой лингвистическую неотличимость, при которой не только обычные разговоры, но даже и дискуссии на темы философии сознания нельзя отличить – как будто они ведутся не между зомби, а между людьми.

Следует отметить, что явный подход к определению понятия зомби страдает концептуальной нечёткостью. Как можно помыслить возможность бессознательного существа, которое поведенчески, функционально и даже физически неотлично от сознательного? На первый взгляд, здесь присутствует противоречие. Поэтому ряд исследователей, например Э. Бейл, считает, что понятие зомби – это не единичное понятие, а тонкие вариации в конструкции *мысленных экспериментов* с зомби, и эти вариации могут иметь важные последствия для тех или иных философских заключений.

Мысленные эксперименты с философскими зомби

Рассмотрим наиболее крупные мысленные эксперименты в хронологическом порядке, который, как считает ряд авторов, отражает собственно *эволюцию проблематики философских зомби* (Т. Полджер, Р. Кирк). Первые два мысленных эксперимента с зомби были предложены Р. Кирком в 1974 г. в двух небольших работах.

«Дэн» **Р. Кирка** (1974 г.). Первый сценарий Р. Кирка выполнен в стиле голливудских зомби из серии фильмов «Я – зомби». В этих фильмах персонаж рассказывает от первого лица, как теряется осознание им окружающей действительности после укуса зомби. В работе «Ощущение и поведение» [Kirk, 1974-1] Кирк описывает отчёты, сообщаемые воображаемым персонажем по имени Дэн. Заразившись неизвестной инфекционной болезнью, Дэн проходит частичные этапы «зомбификации». Вначале он теряет чувство боли. Как-то порезавшись ножом, Дэн не почувствовал боли. Снова воткнув в себя нож, он вскрикнул «Ой», потекла кровь, но никакой боли не ощутил. Он стал себя резать, бить, кричать «а-а-а!», вздрагивать, ругаться, хвататься за раненую часть тела и т.д., т.е. вести себя так, как действовал бы в случае ранения. Однако боли по-прежнему нет. Факты нормального болевого поведения и ненормального отсутствия феномена боли ставят Дэна в тупик. Он идёт в больницу. В процессе тестирования врач его изрядно избивает, однако ставит диагноз – ничего подозрительного. Томография показывает, что не только поведение и внешнее состояние организма, но и мозг Дэна совершенно правильно отражают болевую ситуацию. Однако боль не ощущается! Проходит время. Мало-помалу притупляются другие чувства. На каждой стадии угасания того или иного ощущения Дэн смутно беспокоится по поводу утраты и потери сознательной осведомлённости об утрате. Однако вскоре беспокойство проходит, он забывает о прежних чувствах, приспособляясь к новым условиям с меньшим количеством разнообразия чувственных модальностей. Наконец феноменальная осведомлённость Дэна об окружающем исчезает полностью. Через некоторое время после исчезновения последнего чувства Дэн адаптируется к новой «жизни», уже не осознавая самого факта приспособления. Для всех посторонних он по-прежнему выглядит совершенно нормальным. Однако Дэн стал зомби.

Таким образом, Р. Кирк считает, что мы можем связно вообразить следующее: 1) как Дэн последовательно теряет чувственные модальности, сохраняя при этом поведенческие и физические способности; 2) как Дэн переходит границу абсолютного неосознавания окружающей действительности.

«Гулливер» Р. Кирка (1974 г.). Над вторым сценарием Р.Кирка нависла мрачная тень гаитянских зомби. В работе «Зомби против материалистов» [Kirk, 1974-1] он описывает следующий эксперимент: лилипуты атаковали Гулливера и захватили его голову. Получив доступ к нервной системе Гулливера, они отсоединили его нервы, стали отслеживать входную информацию, поступающую от афферентных нервов, и посылать выходную информацию к эфферентным нервам. Поведение Гулливера при этом ничем не отличается от его поведения в исходном состоянии. Ему присущи все поведенческие диспозиции бывшего Гулливера, однако он ничего не ощущает, не чувствует, не владеет никаким феноменальным опытом. Гулливер стал зомби. Далее сценарий усложняется – голова Гулливера возвращается к исходному состоянию, лилипуты подключают нервы куда следует и покидают «поле боя». Возникает вопрос – станет ли Гулливер ощущающим существом? По-видимому, нет, и сомнительно, что чувства вернутся при восстановлении исходного состояния, ведь нам не известны причины, закономерно связывающие факты сознания с физическими фактами. Здесь возникает проблема определения границы и момента, когда из физических феноменов следуют психологические явления.

Мы видим, что и первый и второй эксперименты Р. Кирка демонстрируют мыслимость зомби. Заключение следующее: сложно представить, как какую-нибудь вразумительную версию материализма можно примирить с логической возможностью зомби при условии того, что мы, люди, являемся сознательными.

«Зимбо» Д. Деннетта (1991 г.). Проблема возможности зомби для Д. Деннетта вовсе не существует. Более того, утверждение – «человек это зомби» выступает в роли базовой посылки оригинального гетерофеноменологического метода, суть которого состоит в объяснении сознания из интересубъективной перспективы третьего лица, в отличие от традиционных, аутофеноменологическо-

интроспективных методов. В работе «Сознание объяснённое» [Dennett, 1991] Д. Деннетт вначале предлагает представить «обычного» зомби – биологическое или искусственное существо, которое ведёт себя сообразно с неосознаваемыми потребностями. Далее он предлагает вообразить более сложную разновидность зомби, который, помимо того, что осуществляет простой зомби, контролирует собственные действия, рефлексируя над ними. Рефлексирующее зомби получает имя *зимбо*. Это – зомби с метапрограммой самоконтроля, благодаря которой можно, не осознавая, управлять информационными состояниями низкого порядка [Ibid, P. 310]. Деннетт предлагает побеседовать с зимбо. Так как зимбо контролирует все свои действия, включая речевые акты, то может ответить на любые вопросы об образах, мечтах, чувствах, верованиях и пр. Ответы зимбо будут казаться собеседнику весьма естественными. Спрашивающий предполагает, что зимбо столь же сознателен, насколько сознателен он. Более того, зимбо сам себя может считать сознательным, так как «...подсознательно полагал бы, что пребывает в ментальных состояниях, определения которых фиксируются задаваемыми вопросами. Зимбо «думал» бы, что обладает сознанием, даже если это не так!» [Ibid, P. 311]. Отсюда – девиз Д. Деннетта: «Все мы – зомби!». Мы, люди – сложные самоуправляющиеся зимбо – можем говорить об образах, мечтах и чувствах, восхищаться, верить, надеяться, любить и пр. Однако если мы полагаем, что сознание – нечто отделённое от наших рассуждений об этом, то глубоко ошибаемся. Мы видим, что проблема *возможности зомби* совершенно не стоит. Сознание, которым владеет зимбо – это нечто вроде здоровья, оно или есть или его нет. Для Д. Деннетта важно другое – как правильно охарактеризовать, вообразить, оговорить образ зомби. Неправильный образ зомби формируется из-за убеждённости в том, что сознание – нечто особое, существующее вне поведения и обозначаемое терминами «квалиа», «интроспективное», «феноменальное» и пр. Однако наше сознательное поведение явно не отличается от бессознательного. Загадка не в том, почему мы что-то осознаём, а почему (и как) сознание при необходимости возникает в существах, поведенчески неотличимых от нас, людей. Любое же дуалистическое упоминание о сознании как о некотором «внутреннем состоянии», узакониваемым самотождественным *Я* в контексте проблемы зомби, приводит, по мнению Д.Деннетта, к «невообразимой

нелепости зомби» [Dennett, 1995-1]. А источник «глупости» – это просто-напросто недостаток творческого воображения у философов. По крайней мере, именно это Д. Деннетт отрицает у авторов следующих двух экспериментов.

«Кремниевые мозги» Дж. Серля (1992 г.). Дж. Серль, раскрывая проблематику зомби в работе «Открывая сознание заново» [Searle, 1992; Серль, 2002], по сути, воспроизводит первый эксперимент Р. Кирка (1974 г.). Однако здесь мы видим компьютерные новации: если у «Дэна» чувственность исчезает в силу неизвестной инфекции, то у Серля – из-за замены компонентов мозга компьютерными чипами. Мозговая модификация ведёт к трем возможным последствиям: V1) ментальная жизнь остаётся прежней, так как чипы дублируют не только вход-выход, но и ментальные феномены, ответственные за функции входа-выхода; V2) ментальная жизнь потихоньку затихает и в конце концов сознание полностью покидает тело, хотя внешне наблюдаемое поведение остаётся прежним; V3) ментальная жизнь остаётся прежней, но сознание никоим образом не влияет на поведение, субъект парализован, хотя продолжает ясно и отчётливо осознавать окружающее.

Возможность «кремниевого» зомби (V2) для Дж. Серля вне сомнения, хотя он спешит с уточнением: имеется в виду не эмпирическая возможность, так как «этот мысленный эксперимент остаётся правомерным лишь в качестве логической или концептуальной возможности» [Searle, 1992, P.78]. Цель зомбиевого сценария состоит в раскрытии логической возможности сохранения прежнего поведения при нарушении нормальных посредствующих отношений между сознанием и поведением. Кремниевые чипы не дублируют каузальные силы мозга, производящие сознательные ментальные состояния, они лишь копируют определённые мозговые функции «вход-выход». Здесь отсутствует ментальная жизнь, возникающая в силу реальных биологических детерминаций между поведением и сознанием [Searle, 1980; Серль, 2006]. Подчеркивая это положение, вспомним «Китайскую комнату», где Дж. Серль утверждает, что именно по причине отсутствия у компьютерных систем биологической каузальности психологических феноменов сильный искусственный интеллект не возможен.

На сценарий Дж. Серля резко реагирует Д. Деннетт: «...конечно, V2 – это логически возможная интерпретация... Но есть и иная, более

жестокая трактовка V2: когда вы [на время] умираете, другое сознание наследует ваше тело. Высказывания, произносимые вашим телом, которые, возможно, вы с трудом улавливаете, – это не ваши высказывания. Но они не могут быть ничьими высказываниями!... Меня обескураживает то, как Дж. Серль раскрыл столь зияющую брешь в амбразуре своего мысленного эксперимента» [Dennett, 1995-2].

Ещё более резко Д. Деннетт реагирует на образы зомби, рисуемые Т. Моуди, О. Фланаганом и Т. Полджером. К шквалу его эмоциональных реакций следует относиться осторожно, здесь имеются разные плоскости изучения тематики зомби. Д. Деннетт – антизомбист, а его оппоненты – нейтральные зомбисты.

Принцип несущественности сознания О. Фланагана (1991 г.). О. Фланаган предлагает не мысленный эксперимент с зомби, а *теоретический принцип несущественности сознания* (сознательного инэссенциализма, «conscious inessentialism»). Принцип формулируется следующим образом: «любая ментальная деятельность М в любой когнитивной области D может осуществляться бессознательно, не дополняемая и не сопровождаемая сознанием, несмотря на то, что мы, люди, обычно осуществляем М сознательно» [Flanagan, 1991, P. 309]. Принцип предполагает, что любая наука о человеке должна решать вопрос – насколько сознание необходимо для организации поведения? Не только интеллектуальное, но и высокоморальное поведение может быть бессознательным, хотя ему обычно приписывают сознательность. Принцип несущественности сознания (ПНС) апеллирует к вычислительному функционализму и к положениям когнитивной науки, отождествляющим информационные и ментальные процессы. Девиз О. Фланагана таков: *разуму сознание не нужно*.

Данный принцип – чисто теоретического плана. По сути, он воспроизводит старую методологическую установку бихевиористской методологии изгнания психологических терминов из теоретических рассуждений о психике. Этим как бы достигается объективность психологии. Сегодня ПНС видится крайне важным не только для методологии когнитивной науки, но и для мировоззрения, так как усиливает скепсис вопросов о роли, адаптивной и эволюционной значимости сознания – ведь сознание не необходимо ни в метафизическом, ни в логическом отношении. И хотя О. Фланаган всячески отгораживается от проблемы «философских зомби» (называя их «болванами»), из контекста очевидна

его зомбистская ориентация. Так, он пишет, что сознание не необходимо в возможных мирах, где обитают функционально эквивалентные нам, людям, существа. Однако в нашем, реальном мире сознание – суть реальность. Почему? Какое эволюционное преимущество получает существо, которому даровано сознание в отличие от бессознательного существа, которое, кстати, может быть чрезвычайно интеллектуальным. Может, сознание ничего и не даёт человеку? [Flanagan, Polger, 1995; 1998], [Фланаган, Полджер, 2006].

Противоречивая позиция О. Фланагана – это либо стремление к строгому теоретическому принципу, либо к метафоре «бессознательных существ, ведущих себя сознательно». Такая метафорфоза представляется типовой для зомбифилов: они плывут между Сциллой необходимости логической строгости понятий и Харибдой метафорико-поэтического образа зомби.

«Земля зомби» Т. Моуди (1995 г.). Т. Моуди, критикуя ПНС в небольшой работе «Общение с зомби» [Moody, 1995], особо выделяет проблему *другого* – можно ли отличить (дискриминировать) сознательных существ от зомби и каковы должны быть научные принципы такой дискриминации. Приводится ряд аргументов в пользу того, что идентификация зомби должна осуществляться не за счёт поведенческих отличий, а на социокультурном уровне, по крайней мере, в условиях речевой коммуникации. Ни логическая, ни физическая возможность зомби даже не обсуждаются. Главная задача – найти критерий «зомбиевости» (mark of zombihood) исследуемой системы – *как отличить личностное от зомбиевого?*

Для определённости критерия «зомбиевости» вводятся следующие классы зомби.

1. «Зомби среди нас» – это те, кто растёт и живёт среди людей. Теоретического интереса такие зомби не представляют – общим местом является тот факт, что многие действия человек выполняет бессознательно.

2. «Истинные зомби» – это существа, изолированные от мира, в котором обнаруживается сознание. Эволюция зомби имеет чисто физические основания и совершенно не зависит от взаимоотношений с сознательными существами – ведь таковых нет в зомбиевом окружении. Для «чистоты» эксперимента Т. Моуди предлагает возможный мир, планету «Земля зомби». Жители её бессознательны. Эволюция на Земле

зомби осуществляется, однако отбору подвергаются исключительно поведенческие акты, эффективные по тем или физическим критериям. В борьбе за «выживание» не участвуют сознательные регулятивы поведения. Их нет, потому что они ничего не значат.

Допустим, между людьми и зомби начался «культурный» обмен: люди стали прилетать на Землю зомби, а зомби – на Землю людей. Осуществление «межкультурного общения», при эксплицировании лингвистических различий разных «культур», создаст достаточные условия выявления зомбиевости. Для этого нужна процедура понимания людьми высказываний зомби и, наоборот, понимания зомби того, что высказывают люди. Для различения человеческих и зомбиевых выражений предлагается индексировать термины, например «понимать[z]» означает аналог слова «понимать» на русском языке. Но что значит для зомби слово «понимать»[z]? Здесь Т. Моуди ссылается на проблему понимания, поднятую Дж. Серлем в «Китайской комнате». Допустим, входные и выходные данные и у зомби и у людей тождественны. Однако человек обладает определённой разновидностью сознательного опыта – для человеческого «понимания» присуща осознанность, осмысленность информации не только на входе и выходе, но и в процессе внутренней, рефлексивной обработки. Для зомби феноменологический опыт отсутствует, термин «понимание[z]» ничего не значит, это слово вообще не могло бы возникнуть на Земле зомби. Зомби, как бы они ни развивались в своём замкнутом физическими законами мире, не способны создать словарь ментальных терминов, включающий слова думать[z], воображать[z], мечтать[z], верить[z], видеть сны[z] и пр. «Философский язык» зомби (Т. Моуди считает, что философия у зомби будет иметь место) также будет различаться – для зомби не будет проблем «инвертированного спектра», «квалиа», «другого сознания» и пр. Они смогут рассуждать[z] по проблеме «других зомби», но эта проблема будет выявлена лишь после встречи зомби с нами. Более того, после встречи с нами зомби станут рассуждать[z] о том, что именно люди говорят о сознании. Возникнет понимание[z] ментальных терминов, высказываемых людьми, по аналогии с тем, как люди понимают высказывания мистиков и эзотериков.

Таким образом, отсутствие у «истинных» зомби словаря ментальных терминов позволит, как считает Т. Моуди, отличить зомбиевое от человеческого. Однако коммуникация между зомби и людьми приведёт в

конечном счёте к тому, что истинные зомби станут «зомби среди нас» и их невозможно будет распознать в нашем, человеческом мире.

Критика «Земли зомби» (с 1995 г.). Мысленный эксперимент Т. Моуди вызвал оживлённую дискуссию. Состоялся специальный симпозиум, представленный в 1995 г. на страницах «Журнала изучения сознания»¹. Название публикаций и имена участников дискуссии говорят сами за себя: Родни М. Дж. Коттерилл – «О единстве сознательного опыта»; К. Сазерленд – «Земля зомби»; О. Фланаган и Т. Полджер – «Зомби и роль сознания»; Д. Деннетт – «Невообразимая нелепость зомби»; Г. Гузельдер – «Многообразие зомбиевости»; Дж. Ленайер – «Зомби обсуждать невозможно»; Д. Маккарти – «Зомби Тодда Моуди»; С. Брингсфорд – «В защиту непостижимости зомби»; М. Мидглей – «Зомби и тест Тьюринга»; А. Элитзар – «Сознание нельзя игнорировать»; К. Чандлер – «Декарт, Фланаган и Моуди»; Д. Ходгсон – «Что зомби не могут делать?»; Ч. Т. Тард – «Да, мы – зомби, но можем стать сознательными»; Т. Моуди – «Почему зомби не хотят умирать?».

Дискуссия продолжается по сей день. Выделим четыре класса отзывов об этом мысленном эксперименте.

1. Некорректность мысленного эксперимента.

Д. Деннетт считает, что этот мысленный эксперимент недостоин рациональной аргументации и представляет собой разновидность «философской глупости»: «Земля зомби» – невнятное самоопровержение собственной позиции. У философов это встречается редко, но в данном случае имеет место. Т. Моуди определяет зомби как существ, поведенчески неотличимых от сознательных людей, поэтому они и в самом деле должны быть неотличимы! Если словарь зомби и философия зомби отличаются от человеческого словаря и человеческой философии, то они на самом деле отличаются! Какая невообразимая нелепость! Здесь итог всего того, что является ложным в современных теориях сознания. Но, упорствуя в своём выборе, философы бросаются понятием зомби так же, как с ладони на ладонь перекидывают горячую картошку.

¹ Symposium on "Conversations with zombies" (1995) Journal of Consciousness Studies. Volume 2, No. 4 (1995), http://www.imprint.co.uk/jcs_2_4.html

П.Чёрчленд более безжалостна: дискуссии по теме «Земля зомби» беспочвенны, а мысленный эксперимент Т. Моуди – это аргумент, демонстрирующий слабость собственно мысленных экспериментов.

С. Брингсйорд считает, что зомби-базированный аргумент против ПНС, на первый взгляд, вполне приемлем. Однако детальный анализ показывает иное: «Земля зомби» не более чем фокус. Составить ясное понятие зомби невозможно, в самой идее зомби скрывается то, что превосходит способность человеческого понимания – объясните нам, как можно сознательно представить бессознательное состояние? Поэтому идентификация зомби сродни пророчеству.

2. Продуктивность идеи Земли зомби для изучения сознания.

Родни М. Дж. Коттерилл считает, что «Земля зомби» ставит новые акценты в проблематике теории сознания – но не на отношении «сознание / мозг», а на связи «сознание / поведение». Сознание следует ассоциировать не со стимулом и последующим восприятием ментального образа (что характерно для первого отношения), а с движением и реакцией. «Сайт» сознания должен быть: а) виртуальной предпосылкой единства сенсорной сингулярности; б) нейтральным относительно высокоуровневых иерархий реакций; в) доступным для реальных отношений обратной связи.

А. Элитзар считает, что «Земля зомби» представляет особый теоретический интерес. «Дух» зомби вызывается для демонстрации уникальности сознания и, несомненно, привлекателен, так как оспаривает исключительность прав физической реальности. Он отказывает в правах теории сознания, которая оберегает полноту и замкнутость физического мира.

3. «Земля зомби» в контексте философии искусственного интеллекта.

Дж.Маккарти – «отец» искусственного интеллекта – считает «Землю зомби» совершенно неправдоподобным экспериментом. С позиции искусственного интеллекта, как полагает Дж. Маккарти, сознание – это сложная совокупность взаимодействующих материальных процессов, а не некая идеалистическая конструкция или дуалистическая субстанция. Сознание играет конкретную роль в человеческой деятельности. Чтобы зомби могли вести себя как люди, им требуется что-то вроде сознания – так называемое «псевдосознание». Псевдосознание должно участвовать в выполнении моторно-двигательных, речевых,

интеллектуальных актов. Сюда относятся и лингвистическая компетентность, по крайней мере, способность зомби отвечать на задаваемые вопросы. Однако, по определению, ничто похожего на псевдосознание у зомби не должно быть, а так как псевдосознательная деятельность неотделима от поведенческого акта, то выносится вердикт: «*Зомби Моуди – плод выдумки Моуди. Зомби Моуди невозможны!*»

А. Сломан – видный философ искусственного интеллекта – в принципе соглашается с Дж. Маккарти. Он исходит из анализа различных форм робототехнической реализации феномена субъективности (позиции первого лица): «[Зомби] не сможет имитировать поведение сознательного существа, не обладая моделями «эмоций», «точек зрения» или иных феноменов субъективной жизни. Утверждающий подобное занимается самообманом. Таким «философам» следует поучиться инженерному делу, тогда они поймут, что робот-зомби работать не будет. Для инженера представить себе робота-зомби означает отказаться от всего того, что он знает о механизмах обработки воспринимаемой информации, хранении её, генерации новых задач, принятия решений для реализации поведения или воздержания от недопустимых действий, оценки степени выполнения задачи или мобилизации ресурсов» [Sloman, 1998-1; Сломэн, 2006].

О. Флаганан и Т. Полджер, несомненно, осуществили наиболее многогранную критику «Земли зомби» [Flanagan, Polger, 1995]. Пристальная критика понятна – ведь ключевая идея Т. Моуди заключалась в том, чтобы показать, что ПНС ложен или нуждается в пересмотре. Авторы основательно разбирают *аргумент Моуди*, который переформулировали следующим образом: а) если ПНС верен, тогда нет возможности отличить зомби от сознательного существа; б) можно утверждать, что жители Земли зомби – это зомби; в) следовательно, ПНС неверен или требует уточнения. Демонстрация аргумента корректна (*modus tollens*), но посылки не обоснованы. Первая посылка ложна, так как ПНС утверждает лишь то, что интеллектуальный акт может быть выполнен без осознания этого акта, но ничего не говорит о ситуациях различения людей/зомби. Вторая посылка ложна, так как легко можно вообразить сценарий, в котором зомби (по сути, робот) вступая в социальные связи (например, для предупреждения других зомби об опасности), формирует квазиментальные термины. Такие же термины возникают в результате сбоя оборудования или формирования непредусмотренных алгоритмов

сценариев действий. Авторы рисуют яркие образы компьютерных реализаций зомби, которые способны эволюционировать, создавать зомбиевый язык, включающий термины самого широкого диапазона значений – от образов, снов и мечтаний до философских споров на темы инвертированного спектра, квалиа, проблемы *другого*.

4. Иерархия сознаний.

Особый «религиоведческий» интерес вызывают утверждения Т. Моуди о разнопорядковых уровнях понимания реальности мистиками и обычными людьми. На эти заявления живо откликается представитель трансперсональной психологии Ч. Тард, который в противовес принципу несущественности сознания предлагает *принцип расширения сознания*: «Все мы – зомби, но можем стать сознательными!». Для трансперсоналистов преодолеть зомбированность – это принять психоделические средства. И тогда «Ваше сознание станет абсолютным».

Дискуссии по «Земле зомби» продолжают по сей день, так как затрагивают важные вопросы методологии искусственного интеллекта, социокультурные параметры соотношения сознательного и бессознательного, но не в традиционном психоаналитическом контексте, а в контексте идей когнитивно-компьютерных наук и искусственного интеллекта. На мой взгляд, ПНС, характеризуя теоретический уровень изучения роли и функций сознания, вполне совместим с мысленным экспериментом земли Зомби, который характеризует риторико-метафорический контекст изучения проблематики зомби.

«Чалмерс-зомби» Д. Чалмерса (1996 г.). Если предыдущий эксперимент служит решению конструктивных вопросов, то Д. Чалмерс продолжает спекулятивную линию проблематики зомби, значительно дополняя её модально-логическими исследованиями. Он предлагает вообразить свой собственный дубликат, искусственную «версию» Чалмерса, организованную точно так же, как организован реальный философ – Чалмерс [Chalmers, 1993-1]. Разница в следующем: там, где у реального Чалмерса нейроны, у «двойника» – кремниевые чипы. Для самого Чалмерса и, как он считает, для многих других очевидно, что «Чалмерс-зомби» не обладает сознанием, ведь у него внутри всё *пусто и темно*, так как ни в кремнии, ни в биохимии нет ничего того, что относится к сознанию. Для Чалмерса также очевидна мыслимость и, следовательно, логическая возможность Чалмерса-зомби. Комментаторы обычно цитируют следующее: «Признаюсь, что

логическая возможность зомби кажется совершенно очевидной для меня... Нет противоречия в данной дескрипции, хотя принятие её логической возможности основано на интуиции... Мне кажется, что едва ли не каждый способен помыслить эту возможность... Я не могу обнаружить никакой логической бессвязности [в идее зомби] и имею ясную картину, когда представляю зомби». Некоторые могут отрицать возможность зомби, однако они должны быть компетентны в проблематике возможности, причём данная компетентность должна быть выше, нежели чем у тех, кто допускает возможность зомби. Короче, доказывать [возможность/невозможность] зомби должен тот, кто утверждает, что дескрипция зомби логически невозможна. При этом [оппонент] должен чётко показать – где видится явное или неявно выявляемое противоречие [Ibid, P. 96, 99]. То есть бремя доказательства лежит на антизомбисте, а не на зомбисте.

Эксперимент с зомби важен для Д. Чалмерса. Формально на это указывает глоссарий книги: слово «зомби» – наиболее часто употребляемое в книге. Чалмерский выбор «логической возможности» (или мыслимости) зомби – это намерение снабдить ярким экстравагантным примером и усилить базовую неодуалистическую посылку о том, что сознательные состояния, или «квалиа», неподвластны физикальному и функциональному анализу.

Однако, как считают критики (Р. Кирк, Э. Коттрелл), Д. Чалмерс слишком уж просто «расправился» с проблемой мыслимости зомби, приступив сразу к аргументу зомби. Также он не последователен в применении аргумента мыслимости зомби. Поэтому критики предлагают Чалмерса-зомби для оппозиции, т.е. для реализации стратегии антизомбистов: если можно убедительно показать *немыслимости* Чалмерса-зомби, тогда и теория Д. Чалмерса рухнет.

Продуктивность идеи зомби

Автор полагал, что, представив ряд явный и неявных определений зомби (конечно, неполный), приуменьшил дозу антизомбиевой вакцины «неведения» от очередной зарубежно-философской «инфекции». Однако здесь его поджидал удар. В последней крупной работе «Зомби и сознание» (ноябрь 2005 г.) Р. Кирк – «отец» проблематики зомби – неожиданно для всех делает обескураживающее заявление: «Книга преследует две основные цели, одна из которых – стереть идею

зомби раз и навсегда!» [Kirk, 2005]. Зомби не стоит заниматься, так как благодаря им возникает *фундаментальная путаница в понятии сознания*. Возможно, для когнитивной науки в целом это так. Однако для контекста э-культуры, ИКТ, когнитивных технологий тест зомби крайне важен – он позволяет выносить экзистенциальные суждения по поводу атрибуции и предикации психических феноменов некоторой х-системе – т.е. суждений типа «система сознательна», «система бессознательна». Перспективы применения теста зомби представлены далее при обсуждении изучения персональных феноменов в контексте э-культуры. До рассмотрения проекта искусственной личности (см. подразд. 5.2) сделаем следующее заявление:

Проблема философских зомби значима по причине двойственности проблеме искусственной личности

Однако подчеркнем следующее:

Для нейтрально-зомбистской позиции когнитивно-компьютерные проекты философских зомби немыслимы.

4.3. Дефиниция ментальных феноменов: «сознание» в электронной культуре

Автору известны две главные директивы философского изучения феномена сознания – метафизико-аналитический и метафизико-синтетический (спекулятивный). В контексте э-культуры второй путь следует напрочь исключить. Иначе наукоёмкие технологии э-культуры покроются шелухой несовместимых суждений из гегельянских диалектических «переходов количества в качество», поэтических хайдеггеровских суждений типа «время на то и время, что оно себя временит», бес-связных эпатажных постмодернистских квазиэкивоков.

Более продуктивная директива, на которую, кстати, опирается современный компьютеринг, возникла в контексте аналитической философии сознания. Она обеспечивает конструктивный и в определённой степени детальный анализ ментальных свойств, качеств и способностей. Анализ используется не ради «анализа». Результаты философской работы позволяют упорядочить, систематизировать, преобразовать, операционализировать, интегрировать компоненты анализа, представ-

ленные в формате верифицированных или фальсифицированных языковых выражений для различного рода практических целей – оптимизации экономики, эффективности политики, экспертизы произведений искусства, достоверности истории, рационализации права и пр. В нашем случае цель философского анализа состоит в изучении возможности технологической реализации феномена сознания.

По причине собственных методологических предпосылок детального и скрупулёзного изучения предметной области, проведения специальных исследований в области языка описания предметной области и языка описания метода описания и пр. метафизико-аналитическая стратегия «грешит» объёмностью и многочисленностью научных работ. Так, Д. Чалмерсом насчитывается 28 246 крупных книг и статей [Chalmers Bourget, 2013] по проблеме сознания на момент написания автором данного текста (17.05.2010 г.). Конечно, этот список значительно больше. Поэтому крайне затруднительно, а по сути невозможно, разобраться в трудно регистрируемом многообразии подходов, парадигм, теорий, концепций, мысленных экспериментов, эмпирических данных и пр. Вполне естественно стремление исследователей ввести порядок в это разнообразие. Поэтому Д. Чалмерс создаёт иерархическую классификацию работ по проблеме сознания и вводит систему кодирования (очень примитивную с позиции библиотечного дела). Само собой разумеется, что данная классификация служит не энциклопедическим целям, а конкретным задачам, стоящим перед исследователем. Имеются иные модели.

Продуктивной концептуальной моделью исследования проблемы сознания представляется «философский квадрат» (рис. 2, а), предложенный Д.И. Дубровским более 20 лет назад (воспроизведён в работе [Дубровский, 2007, С. 26 – 27]). Модель проста, однако надёжно систематизирует нашу работу. Модель часто применяется в работах Д.И. Дубровского и его учеников. Так, в одной из последних работ он применил её к проблеме интересубъективности в контексте исследований ИИ [Дубровский, 2011-1, С. 129 – 148].

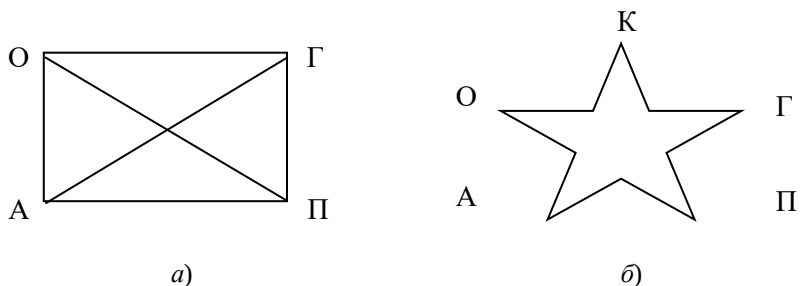


Рис 2. Концептуальная модель проектов философии сознания: *а* – «философский квадрат» Дубровского; *б* – культурологически нагруженная пентаграмма проектов философии сознания

Четырёхмерное пространство схемы исследований, предложенное Д.И. Дубровским, определяется четырьмя фундаментальными категориями: 1) онтологическое (О), 2) гносеологическое (Г), 3) аксиологическое (А), 4) праксеологическое (П). Так как эти категории, по его мнению, инвариантны по отношению ко всем философским направлениям, то они соответственно конституируют и частную проблематику – проблему сознания (точнее, субъективной реальности) [Дубровский, 2009-2]: «Без основательной гносеологической рефлексии невозможна основательная онтология... Ценностные параметры и факторы активности непременно включены в любой акт познания и сказываются на его результате» [Там же, С.26]. Отношения, обозначенные на «квадрате», позволяют вести систематический и целостный учёт параметров сознания, не упускать из виду одно по причине изучения другого. Представленные на «квадрате» планы изучения сознания, в силу философского обобщения, позволяют высказывать универсальные суждения. Однако понятие «сознания» релятивно просто по причине его феноменального, необъективируемого статуса, уникальности индивидуального сознания. Конвенции по поводу того, что считать «сознанием», предзаданы социокультурными контекстами употребления данного термина. Поэтому «философский квадрат» не достаточен для отражения фундаментальных отношений, возникающих при познании сознания. Он не учитывает социокультурную обусловленность феномена сознания. «Квадрат» не фиксирует *культурологического измерения* (К). Конструкцию следует дополнить до пентаграммы (рис. 2, б).

Покажем необходимость культурологической параметризации проблемы сознания путем сравнения того, что понимается под «сознанием» в западной и восточной традициях.

Так, например, Ван Гулик – представитель западной культуры – фиксирует не более десятка понятий «сознания» [Gulick, 2004]. Наиболее значимые параметры сознания – это ощущение, бодрствование, самоосведомлённость, субъективность, транзитивность. Основные состояния сознания: «осведомлённость о», квалиа, феноменальные состояния, состояния проекции (что-значит-быть-как), акцессное состояния (состояние доступа), нарративность. Проблема сознания прорисовывается в постановке ряда вопросов:

1) дескриптивные вопросы изучения сознания (каковы особенности сознания?): «данные» от первого, второго, третьего лица, квалиа, феноменальная структура, субъективность, самоперспектива, тождество сознания, интенциональность и прозрачность, сознание как «поток»;

2) объяснительные вопросы (как сознание существует?): редуктивные и нередуктивные программы объяснения феномена сознания и, конечно, наличие объяснительного провала в объяснении;

3) функциональные вопросы (зачем сознание?): каузальный статус для организации поведения, социальной координации, адаптации и контроля; функциональный статус репрезентаций, доступ к состояниям сознания (внутренняя мотивация и свобода воли также относятся к функциональной составляющей проблематики сознания).

Теперь учтём восточную традицию и рассмотрим понятие сознания в психологии раннего буддизма, описанное ламой Анагарика Говинда (Психология раннего буддизма. СПб.: Издательство «Андреев и сыновья», 1993). Здесь насчитывается 54 класса чувственных состояний сознания, большинство которых «эстетически нагружено». Среди них – радость, грусть, боль. Остальные классы (общее число классов состояний сознания достигает 121) – это «этически» нагруженные состояния сознания. По всей видимости, подобного рода детализация, инспирированная метафизикой буддизма, более «аналитична», чем её западный аналог, рассмотренный немного выше.

Данный пример приведён для демонстрации изучения понятия сознания с учётом социокультурных условий. Для целей нашей конкретной работы проблема сознания – это не некая универсальная про-

блема «вообще», но проблема, специфицированная контекстом культуры, в первую очередь, электронной культуры. Причем разные формы э-культуры существенно модифицируют смыслы и значения термина «сознание».

К сожалению, «квадрат» и «пентаграмма» не удовлетворяют стандартным требованиям аналитической философии – в них отсутствуют лингвистические отношения между обозначенными терминами. Необходим психосемиотический контекст.

Двухмерная семантика ментальных терминов

Плодотворной представляется семантико-онтологическая схема психофизической проблемы, предложенная Э. Вильянуэвой [Вильянуэва, 2006, С. 53]. Э. Вильянуэва предлагает плоскую 2-D схему (см. рис. 3), на которой выделяет интенционалы: 1) суждений о феноменальных свойствах сознания; 2) суждений в формате научных объяснений этих свойств.

В целом, исследовательская установка Э. Вильянуэвы гуманистически ориентирована. Он борется с так называемой «парокиальностью» – замкнутостью исследователя в собственном мире субъективных верований, знаний, переживаний, установок. Свобода учёного обеспечивается выходом на просторы объяснительного анализа внутренних психических качеств с использованием всего спектра современных наук – психологии, нейрофизиологии, физики, химии. Однако, убегая от «парокиальности», он вновь попадает в тюрьму парокиальности. Автор – представитель традиционной культуры. Он не задаётся вопросом технологической реализуемости феномена сознания, а для нас это – ключевой вопрос э-культуры. Предпринятые им рассуждения технологического плана (а именно, при рассмотрении функционализма машины Тьюринга) применяются в формате теоретического объяснения психики, но не как инженерный проект. Поэтому схема Э. Вильянуэвы, только появившись, сразу устарела. Она пригодна для проблематики сознания лишь в рамках традиционной культуры.

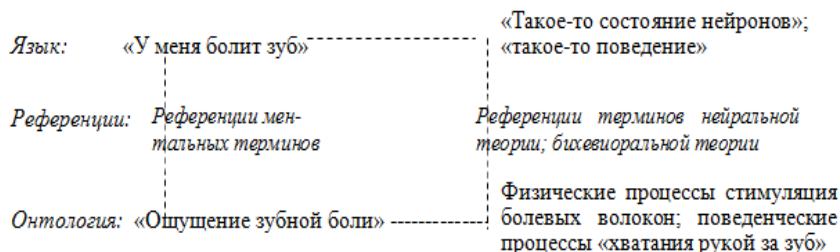


Рис. 3. 2D-семантическая модель изучения проблемы сознания (в традиционной культуре)

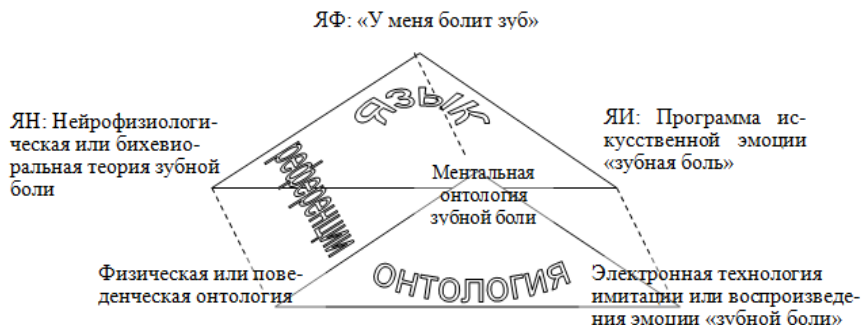


Рис. 4. 3D-семантическая модель изучения проблемы сознания (в э-культуре)

Предложим модель, которая задает технологический императив реализации феномена сознания – трёхмерный концептуальный каркас. Данная модель (см. рис. 4) для традиционной культуры не пригодна, но для выражения специфики э-культуры представляется необходимой – в ней отражается технологическая составляющая реализации когнитивных феноменов и, так же, как в предыдущей – семантико-онтологические связи (референции обозначены пунктирной линией).

Трёхмерная семантика ментальных терминов

Трёхмерная семантика суждений о феномене сознания означает совместное изучение высказываний трёх планов, выраженных на трех специальных языках: ЯФ – язык выражения моих психических состояний, ЯН – язык наук, ЯИ – язык инженерной реализации:

1) ЯФ – *феноменальное суждение*, характеризующее собственно факт сознания, переживаемый человеком (*первичный интенционал*);

2) ЯН – *научное суждение*, рационально объясняющее факт сознания посредством теоретико-эмпирических исследований естественных, гуманитарных, социальных наук (*вторичный интенционал*);

3) ЯИ – *инженерное суждение* о компьютерной реализации феномена сознания, т.е. высказывание в терминах теории компьютерной системы (*третичный интенционал*). Имеется в виду высказывание о структурно-функциональном, логико-математическом, программном, информационном, лингвистическом, методическом и прочем устройстве компьютерной системы.

Возникает правомерный вопрос – зачем применять 3D-семантику, нет ли возможности совместить два плоских плана, один из которых уже предлагался Э. Вильянуэвой? То есть слева оставить квадрат из заданных им семантических соотношений для феноменальных качеств и научных теорий. Справа пристроить подобного рода квадрат, характеризующий референции инженерно-технологической конструкции, которая базируется на той или иной научной теории.

Иначе вопрос звучит следующим образом – «Нуждается ли искусственный интеллект в исследованиях естественного интеллекта?». Эта тема специально обсуждалась на семинаре НСММИ РАН 01.11.2006 г. в докладе О.П. Кузнецова на тему «Может ли искусственный интеллект развиваться независимо от исследований естественного интеллекта?». Оппонентом выступила М.А. Холодная, которая отстаивала необходимость и первоочерёдность исследований естественного интеллекта. Т.Б. Кудряшова высказала идею о том, что исследования ЕИ без исследований ИИ невозможны. Автор высказал идею о совместном, параллельном исследовании. Данная идея ниже получает конструктивное оформление.

Выскажем свою точку зрения. Инженерия ИИ развивается относительно независимо от достижений той или иной научной теории. Са-

молёты стали летать до появления теории полёта, которая так и не состоялась. Универсальная теория интеллекта отсутствует (если вообще таковая возможна). Однако в быту и производстве успешно функционируют компьютерные системы, имитирующие интеллектуальную деятельность человека. Знание нейрофизиологической теории зубной боли не определяет ничего конкретного для инженерной имитации соответствующего поведения, лишь наводит на аналогии. Поэтому здесь важен метафорический контекст исследований сознания для технологической реализации его феноменов. Совершенно лишена смысла буквальная трактовка терминов, скажем, нейрофизиологической теории в терминах организации системы ИИ. Достаточно ясно высказался по этому поводу Г.С. Поспелов: «за спинами специалистов по искусственному интеллекту стоят *тени* великих философов» (курсив мой – А.А.) [Будущее ИИ, 1991].

Здесь были приведены рассуждения исключительно по поводу имитации феноменов сознания, более того, функциональной имитации этих феноменов. Вопрос относительно реального воспроизводства или, собственно, фактической реализации феноменов сознания на базе электронных технологий будет решён лишь в перспективе осуществления широко известного проекта Д.И. Дубровского – проекта расшифровки нейродинамических кодов субъективной реальности. Ничуть не раньше! Автор в этом глубоко убеждён. Также имеется уверенность в том, что предложенная 3D-схема послужит надёжным (по причине простоты) подспорьем для исследования бескрайних просторов теорий сознания с целью поиска технологических решений имитации, моделирования или репродуцирования феномена сознания. Модель требует дополнения, поскольку в ней представлена только одна – онтологическая – составляющая многомерного пространства исследования феномена сознания. Не составляет особого труда проведение аналогичной работы для фиксации гносеологического, аксиологического, праксеологического и культурологического измерений (см. расширение «философского квадрата Дубровского» в нашей «пентаграмме»). После этого следует осуществить требуемые референциальные, интенциональные и экстенциональные привязки к соответствующим языкам, на которых выражаются феномены познания, ценностей, практики, культуры. Это требует кропотливой работы в тех сферах, которые сегодня никак не связаны с технологией: какое отношение, кроме морализаторского, имеет этика к тех-

нологии? В трёхмерном каркасе пенттаграммы эти связи становятся необходимыми. Однако для начала изучим культурологические модификации термина «сознание», соотнесённые с онтологическими типами э-культуры. Смыслы «сознания» в разных формах э-культур, конечно, значительно разнятся.

Смысловые деконструкции «сознания» в электронной культуре

Феномен сознания – неотъемлемый параметр духовной сферы культуры. Вернёмся к онтологической схеме трехчастного деления духовной сферы э-культуры для определения конкретной разновидности «сознания», с которым соотносится технологическая проблема «сознания».

Применим *долевую модель*. Она характеризует долю (степень) участия собственно человеческой сознательной деятельности по «реализации» некоторого когнитивного феномена (ощущения боли, высказывания суждения, создания произведения, осознания самости и пр.) относительно компьютерной системы, которая способна симитировать, смоделировать, репродуцировать эту деятельность (своеобразный квантификатор флангановского принципа несущественности сознания).

Э-культура-1, согласно данной модели, является частью традиционной культуры. Это, например, презентация на DVD предметов народного творчества, видеоролик с выступлением творческого коллектива и пр. Здесь мы не усматриваем особых изменений традиционной проблематики сознания. Человек – полноправный творец э-культуры-1. Для презентации результатов сознательной деятельности, осуществления коммуникаций, поддержки своего творческого процесса (например, компьютерной генерации музыкальных произведений, автоматического реферирования текстов) он пользуется компьютерным инструментарием.

Э-культура-2 это технологически модифицированная форма традиционной культуры. Здесь имеют место превращённые формы сознания, связанные, в первую очередь, с «погружением» субъекта в виртуальную реальность. Аналогом «неэлектронных» способов изменения

сознания являются всевозможные трансперсональные практики.¹ В нашем случае имеют место не медитативные и психоделические средства – стандарт трансперсоналистов, а компьютерное снаряжение систем виртуальной реальности. Человек уже не является абсолютным творцом феномена сознания. Реальность э-культуры-2 наравне с человеком конструирует электронная технология.

Э-культура-3 – самостоятельная форма собственно электронной культуры. Данный случай представляется ключевым для целей нашей работы. Именно в рамках э-культуры-3 наблюдаем максимальную деконструкцию смысла понятия «сознание». В электронном формате представлен и источник, и продукт, и сам способ «сознательной» деятельности. Человек полностью отсутствует, либо его участие ограничивается инициацией электронной технологии и/или получением её продуктов. «Творцами» выступают, например, беспилотные самолёты кибервойск или колонии роботов, «работающие» в условиях, несовместимых с человеческой жизнью².

Проблема сознания в э-культуре-3: к вопросу о «сознании» «чистого искусственного интеллекта»

Дискуссия по поводу наличия «сознания» у «творцов» э-культуры-3 насчитывает порядка 60 лет. Она возникла сразу после выхода статьи А.М.Тьюринга в 1950 г. Собственно данную дискуссию Тьюринг и инициировал, обсуждая возможности идентификации феномена сознания у цифровых вычислительных машин. Споры продолжаются до сих пор. Возникают следующие вопросы относительно проблемы сознания: может ли компьютер ощущать, творить, чувствовать, любить и пр. Например, в 1980 г. Дж. Серль рассматривал такой параметр феномена сознания, как интенциональность, и задавался вопросом «может ли компьютер понимать?». При этом он ввел достаточно грубые понятия сильного и слабого искусственного интеллекта (грубые – с по-

¹ См., например, сборник трансперсоналистов: Пути за пределы «эго» / Под ред. Р. Уолша и Ф. Воон: пер. с англ. М. Папуша, Е. Поле, К. Андреевой. – М.: Открытый мир, 2006. – 392 с.

² См. информацию подобного рода на Интернет-дайджесте «Интеллектуальные Информационные Технологии»; http://www.dialog-21.ru/news/about_digest.asp

зиции возможности наличия более широкого спектра реализации когнитивных феноменов) [Searle, 1980, Серль, 2006]. Сторонники слабого ИИ полагают, что компьютерные программы моделируют частные аспекты ментальной деятельности человека. Их оппоненты – сторонники сильного ИИ – полагают, что соответствующим образом запрограммированный компьютер фактически обладает способностью понимания. Более того, компьютерная программа объясняет способность понимания человека.

Данные дискуссии отражают частные аспекты воспроизводства или имитации феномена сознания электронными технологиями. Для решения фундаментальных вопросов важны рассуждения не о каких-то частных качествах и способностях сознания, т.е. не анализ атрибутивных или предикативных суждений, приписывающих те или иные параметры сознания системам искусственного интеллекта. Необходим анализ *экзистенциального суждения о факте сознания как таковом* – есть «сознание» в системе или «сознание» напрочь отсутствует. Если такое суждение будет иметь место, тогда возможно построение онтологии сознания искусственной системы. Если такое суждение вынести не представится возможным, тогда онтология ментального становится невозможной.

Для отсеивания ненужных дискуссий осуществим радикальный подход с позиции философских зомби – бессознательных систем, которые поведенчески, функционально и/или физически тождественны, неотличимы и/или подобны сознательным существам [Алексеев, 2005-3; 11; 12; 13; Алексеев, 2006-4; Алексеев, 2008-2; Алексеев, 2009-5].

Экзистенциальная атрибуция «сознания» компьютерной системе

Экзистенциальная атрибуция «сознания» – это утверждение факта наличия сознания в х-системе. Лишь после того как вынесено экзистенциальное суждение, появляется возможность качественно и количественно дифференцировать ментальные феномены, приписываемые системе. Несомненно, это радикальный подход. Но он представляется честным и не позволит судить о том, присущ или не присущ «интеллект» системе, если, например, она не обладает «сознанием». Если «сознания» нет, то и «интеллекта» нет. Например, Н.С. Юлина полагает, что машина может мыслить (Н.С. Юлина неоднократно высказывала

это как на конференциях, так и в личных беседах с автором). Но если далее она признает, что интеллектуальная машина сознанием не обладает, то в силу убедительного *принципа несущественности сознания* Фланагана (интеллекту сознание не требуется) мы должны признать, что теория интеллекта должна разрабатываться отдельно от теории сознания. Это не вяжется ни с народно-психологическими представлениями о соотношении мышление/сознание, ни с более рациональными вариантами теорий сознания (кратко было рассмотрено автором в тесте Блока: интеллект – вторичная способность относительно квалиа). Возможны иные варианты. В любом из них тест зомби, используемый в составе КТТ, позволяет четко идентифицировать и структурировать следующие исследовательские позиции.

1. Понятие зомби позволяет навести порядок в определении «сознания», так как предел любому понятию полагается противоречивым или, по крайней мере, противоположным понятием. То есть понятие бессознательного зомби логически необходимо для выработки понятия сознательного существа. Здесь становится важной *таксономия зомби*, учитывающая следующие параметры: а) теории сознания, которые анализируемые зомби опровергают либо удостоверяют (в первую очередь, физикализм, бихевиоризм, функционализм); б) модальности (степени мыслимости и возможности зомби) и др.

2. Зомби дают богатый фактуальный материал в изучении бессознательного. Конечно, эти факты воображаемы, тем не менее, пройдя через горнило мысленных экспериментов, они становятся вполне убедительными. Дозомбиевые средства экспликации бессознательного – это неработоспособные понятия о бессознательном «вообще» либо психоаналитическое эмпирическое мелкотемье «минимальной достаточной определённости» понятия бессознательного, на что справедливо указывает Д.И. Дубровский [Дубровский, 2006].

3. Мысленные эксперименты с зомби аккумулируют, по сути, параметры всех крупных мысленных экспериментов в философии сознания, таких как «Мозги в бочке» Х. Патнэма, «Человек-имитация» Д. Кэмпбелла (1970 г.), «Гомункулная голова» и «Китайская нация» Н. Блока (1978 г.), «Китайская комната» Дж. Серла (1980 г.), «Мэри» Ф. Джексона (1982 г.), «Танцующее квалиа» Д. Чалмерса (1996), «Воришка феноменального» М. Линча (2004 г.) и др. Они раскрывают лишь частные аспекты зависи-

мости сознательных феноменов от физического, бихевиорального, функционального, персонального, социального. Поэтому очевидна методологическо-интегративная функция понятия зомби относительно этих экспериментов.

4. Важность проблематики зомби вызвана не только интересами спекулятивной метафизики сознания, как у Д. Чалмерса. Приложения проблематики зомби можно проследить во всех науках, в поле зрения которых – роль, функции, структуры сознания. Очевидны их применения в общественных и гуманитарных науках. Сегодня зомби находятся в фокусе внимания исследователей искусственного интеллекта и служат основанием для критики либо поддержки таких перспективных направлений в информационной технологии, как искусственная жизнь, искусственная личность, искусственное общество. Способы идентификации зомби сродни тесту Тьюринга – базовому мысленному эксперименту философии искусственного интеллекта, предназначенному для определения интеллектуальности систем. Тест на предмет выявления сознательности/бессознательности систем (*тест Зомби*) – это, вне всяких сомнений, разновидностью теста Тьюринга. Тест Зомби плюс знание внутренних структурно-функциональных отношений тестируемой системы создаёт модифицируемые условия определения сознательности/бессознательности. А это существенно приближает к решению проблемы *другого*.

5. Все вышесказанное подтверждает, что вряд ли Р. Кирк способен убить своё детище – джин выпущен из бутылки. Однако чтобы воспользоваться «силой» зомби, надо отдавать себе отчёт – где мы имеем метафорико-риторические параметры, присущие мысленному эксперименту, а где – логико-эпистемологические параметры, приближающие нас к онтологии. Эти параметры не надо смешивать. Иначе возникает путаница, но не в понятиях философии сознания, а в уровнях описания сознательных систем.

В заключение отметим то, что экзистенциальная дефиниция проблематики компьютерного сознания, ориентирующая исследователя на выявление факта наличия сознания у *x*-системы, является первичной определительной функцией в проблематике приписывания оцениваемой системе ментальных, в частности, и когнитивных феноменов, в общем.

Последующие задачи дефиниции представляются вторичными: атрибутивная, приписывающая психические качества x -системе, и предикативная задача, приписывающая более сложные когнитивные феномены, например феномены социального порядка. Таким образом, решение поднятых проблем «сознания» и *другого* в информационно-коммуникационной среде возможно лишь после зомбистского, анти-зомбистского или нейтрально-зомбистского позиционирования исследователя философии искусственного интеллекта.

Попытка выбора такой позиции приводит к бесконечным схоластическим спорам относительно проблемы «сознания» – можно бесконечно долго комбинировать частные ТТ и манипулировать ими для подтверждения/опровержения собственной позиции по проблеме сознания – такая рефлексия уводит нас от решения психотехнологической проблемы и заводит исследования ИИ в тупик.

Имеются иные стратегии развития ИИ, помимо решения проблемы сознания («сознание-мозг-искусственный интеллект»), как доминирующей проблемы ИИ [Дубровский, 2005]. По мнению автора, стратегию развития ИИ следует формировать исходя из первичного решения более узкой проблемы – проблемы «творчества», но не с проблемы «сознания». В проблематике творчества имеется гораздо больше конструктивного материала; многое имеет чёткий семиотический облик, а не метафизические фантазии об онтологии другого сознания, включая туманную феноменологическую теорию бытия собственного сознания. (Обоснованию переориентации стратегии ИИ с проблематики «сознания» на проблематику «творчества» посвящен подразд. 6.1.)

ГЛАВА 5. КРИТИЧЕСКАЯ ФУНКЦИЯ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

Рассмотрим критический потенциал комплексного теста Тьюринга на примере концептуального анализа инженерных, теоретических и социокультурных проектов э-культуры.

5.1. Критика технологии: комплексный тест Тьюринга против комплексной технологии динамических экспертных систем реального времени

В эпистемологических исследованиях искусственного интеллекта выделяется две фундаментальные оппозиционные парадигмы развития «знаний» компьютерной системы: универсальная и эволюционная.

Оппозиция универсально-эпистемологической и эволюционно-эпистемологической технологий построения интеллектуальных систем

Универсально-эпистемологическая (УЭ) парадигма предполагает полный и постоянный охват предметной области с позиции всезнающего и вездесущего разработчика – эксперта, когнитолога (инженера по «знаниям»), программиста. Для представления декларативных «знаний» такой вариант применим лишь к простым системам с конечным множеством актуально или потенциально обозримых «знаний». Очевидно, что разработчик не в силах специфицировать все возможные ситуации и сценарии сложной предметной области – слишком уж она испещрена уникальными вариациями, не попадающими под общие дефиниции.

Для процедуральных «знаний», генерирующих многообразие предметной области из аксиоматики по правилам вывода, ограниченность универсального подхода менее очевидна. Тем не менее, возможность применения конечного логико-аксиоматического каркаса, реализующего давнюю идею лейбницаанского универсального исчисления, ограничена большими системами, бесконечное множество «знаний» которых потенциально перечислимо. Однако ни для малых, ни для больших систем ИИ не нужен – достаточно традиционных способов обработки информации. Причём принципиально не важно, что такие системы могут быть нечёткими – особого труда не составляет расширение традиционного логико-алгебраического базиса неклассическими, нечёткими, вероятностными, квантовыми и другими вариациями.

Искусственный интеллект востребован для сложных систем, где каждая «регулярность» сопровождается неопределённым веером сингулярностей. Собственно по этой причине для развития «знаний» сложной системы применяют технологии её постоянного и непрерывного сопровождения разработчиком. Такие системы по непонятным причинам называют «интеллектуальными динамическими системами реального времени». Интуитивно очевидно, что к ним невозможно применить термин «искусственный интеллект», ведь здесь задействован исключительно естественный интеллект разработчиков, усиленный компьютерными методами и средствами. Получается, что универсально-эпистемологическая парадигма развития «знаний» априори искажает идеи искусственного интеллекта и сводит их на нет. Тем не менее, она доминирует в современной технологии.

Эволюционно-эпистемологическая (ЭЭ) парадигма адекватна идеям интеллектуальных компьютерных систем. Суть её заключается в приписывании машине способностей самостоятельно, свободно, творчески извлекать и порождать «знания» предметной области, включая «знания о себе». ЭЭ-идеи, концептуально встроенные в механизмы функционирования системы, предполагают активное изменение способностей восприятия и преобразования внешней и внутренней среды, исходящую изнутри системы целеполагающую модификацию форм и способов аргументации логических рассуждений, самостоятельное формирование нормативных регулятивов поведения и деятельности. ЭЭ-концепции в экспертных системах задают автоматическое извлечение фактов предметной области, формирование нетривиальных правил, активное порождение гипотез, креативное продуцирование «знаний». Для компьютерных игр они призваны обеспечить внутреннюю способность расширять пространство сценариев и ролевых статусов. Системы доказательства теорем и автоматизации рассуждений предполагают самоуправляемое расширение логических средств. В системах решения проблем и планирования – формирование мотивационно детерминированных способов целеполагания, смыслового знакового опосредования и, в конечном итоге, «квазиосознавания» окружающей среды и репрезентации её в целостном, независимом, отстранённом формате. В системах технического восприятия, таких как распознавание образов и паттернов, визуальный анализ сцен, слежение за движением объектов и

распознавание лиц, ЭЭ-идеи определяют способы автономного обучения, самостоятельное формирование эталонных образцов, программно-генетическое кодирование решающих правил распознавания.

В системах компьютерной лингвистики, таких как анализ и синтез устной и письменной речи, распознавание отдельных выражений и связной речи, понимание естественного языка и машинный перевод, ЭЭ-концепции предполагают «искусственный семиозис» – самогенерацию знаковых систем, самоорганизацию интегральной системы «синтаксис-семантика-прагматика», автоматическое формирование референций, дескрипций и пр.

В робототехнике ЭЭ-концепции принципиальны для многообразных проектов самовоспроизводящихся, антропоморфных, эмоциональных, сознательных и прочих роботов, особенно если они предназначены для совместной работы в условиях, несовместимых с человеческой жизнью, например на Марсе, Крайнем Севере, в глубинах Мирового океана – в этих проектах колоний роботов задействованы всевозможные модели развития и функционирования общественных форм сознания – от схем предотвращения физических столкновений до способов выработки «квазирелигиозных убеждений».

Представленная панорама применения ЭЭ-моделей имеет в большей мере прогностический статус – такие системы исследователи ИИ называют «нетрадиционными», чтобы подчеркнуть их отличие от общепринятых систем, развивающихся в рамках УЭ-подхода.

Для изучения двух выделенных парадигм воспользуемся методом комплексного теста Тьюринга, который изучает когнитивно-компьютерные феномены на основательном базисе категориально значимых мысленных экспериментов, не растворяясь в спекуляциях теорий сознания и не утопая в эмпирическом мелкотемье бесчисленных проектов интеллектуальных систем.

Эволюционно-эпистемологические идеи в комплексном тесте Тьюринга

Эволюционно-эпистемологические идеи комплексного ТТ выявляются на основе анализа ЭЭ-идей в составе частных тестов, их сопоставления и комбинирования, выделения общих, частных и особенных положений, выдачи рекомендаций относительно реализации ЭЭ-механизмов и пр. [Алексеев, 2012-2]. Рассмотрим в качестве примера как комплексуются ЭЭ-идеи в тьюринговых тестах Д. Деннетта и Дж. Маккарти. Беглый взгляд на комплексный ТТ позволяет судить о том, что тема ЭЭ имманентна проблематике ИИ.

Оригинальный ТТ (тест Тьюринга). ЭЭ-концепция заключается в идее обучаемой машины.

Параноидальный ТТ (тест Колби). Тест служит эталоном для изучения экспертных систем, в которых напрочь отсутствуют ЭЭ-концепции. Такие системы реализуют концепции разработчиков как «навязчивые идеи» и не имеют внутренних механизмов рационального роста «знаний».

Гендерный ТТ (тест Геновой). Тест включает в себя многообразие ЭЭ-вариаций идей гендерного ТТ – от концептуального анализа половой детерминации интеллектуальных способностей до обоснования проекта бесполового постчеловечества.

Креативный ТТ (тест Лавлейс). Тест инициирует тематику компьютерного творчества и опровергает/защищает аргумент Лавлейс, согласно которому компьютер творить не может. Задаёт достаточно строгие аксиоматические критерии продуцирования творческих артефактов. ЭЭ-модели – необходимая составляющая креативных компьютеров, УЭ-модели принципиально не способны реализовать этот тест.

Субкогнитивный тест (тест Френча). В тесте явно звучит тезис: только тот компьютер способен пройти тест, в котором имеются механизмы, позволяющие ему «жить жизнью людей», и возможно прохождение всего пути эволюции человеческого интеллекта либо имитация этого процесса.

Глобальный ТТ (тест Блока). Тест обеспечивает наиболее демонстративное изучение УЭ-концепции. Он задаёт необходимое условие приписывания интеллекта х-системе, с учётом антибихевиористской критики оригинального ТТ. Раскрывается ограниченность лингви-

стического бихевиоризма тьюринговой игры в имитацию и функционализма машины Тьюринга. Новая формулировка ТТ (нео-ТТ) предполагает диспозиционные способности системы продуцировать осмысленную последовательность вербальных реакций на любую последовательность вербальных стимулов. Эти способности поддерживаются машиной Блока – глобальной экспертной системой, которую разрабатывает всё человечество. Внутреннее устройство машины убеждает, что её «интеллект» – это интеллект разработчиков, но никоим образом не её собственный. ЭЭ-идеи связываются с метафизикой психофункционализма.

Китайская комната (тест Серля). Тест отрицает концепцию сильного ИИ, согласно которой при прохождении ТТ компьютер не только моделирует человеческую способность к пониманию (это характерно для позиции слабого ИИ), но фактически «понимает» и объясняет её. Феномен «понимания» представлен в контексте интенционалистской теории сознания, которая обосновывает сильный антикогнитивистский тезис о невозможности «сильного» ИИ. ЭЭ-идеи инспирированы биологическими основаниями феномена интенциональности.

Инвертированный ТТ (тест Ватта). ЭЭ-идея состоит в изучении феномена воображения в вычислительном контексте и формирования рефлексивного «зеркала» в процессе обращённости машины на себя (интеллектуальная машина, тестирующая собственный интеллект).

Тотальный ТТ (тест Харнада). Тест предполагает создание искусственных ЭЭ-механизмов, полностью неотличимых от естественных.

Тест Тьюринга (тест Швайзера). Это – самый тотальный тест, который дополняет предыдущий тест, погружая компьютерную систему – искусственную личность – в контекст эволюционного развития сообщества других таких же систем, по всем своим социокультурным проявлениям неотличимых от людей.

Имеется ряд других интересных версий и интерпретаций ТТ. Подчеркнем, что из всех частных тестов можно явно или неявно реконструировать сюжеты концептуальной драмы между ЭЭ-идеалом интеллектуальной системы и УЭ-возможностями современной компьютерной технологии её построения. При изучении ЭЭ-парадигмы в контексте КТТ возникают следующие вопросы: 1) дефинитный – как по вер-

бальному поведению x -системы определить наличие в ней ЭЭ-приспособлений? 2) конструктивный – как реализовать ЭЭ-механизм и, собственно, возможно ли это? 3) критический – может ли интеллектуальная искусственная система самостоятельно развиваться? 4) мировоззренческий – каковы социокультурные последствия применения ЭЭ-систем?

ЭЭ-парадигма развития «знаний» представляется аподиктическим условием интеллектуальности компьютерной системы – так, по сути, считает большинство философов и теоретиков ИИ, начиная с А. Тьюринга и Дж. Маккарти. Поэтому положительное решение ЭЭ-вопросов служит поводом приписать x -системе интеллект. Если ЭЭ-вопросы не решены, то нечего рассуждать об ИИ.

В рамках данного параграфа ограничимся частичным решением дефинитного вопроса, в общих чертах рассмотрим конструктивный вопрос и оставим без внимания критические и мировоззренческие положения. Выдвинем гипотезу, апеллируя к заявлениям ряда видных исследователей ИИ – обнаружить в x -системе самостоятельные, свободные, творческие компетенции можно тогда, когда в ТТ включены вопросы, на которые ожидаются ответы с позиции «здравого смысла» (Common Sense). Само собой разумеется, что понятие «здравый смысл» крайне расплывчато и всячески изгоняется из современного научного дискурса. Однако крайне частое употребление данного термина в методологии ИИ просто вынуждает к нему прибегнуть. Впервые данное понятие как критерий демаркации интеллектуальной компьютерной системы от неинтеллектуальной применяет основатель ИИ – Дж. Маккарти: *«Здравый смысл», приписываемый компьютеру, это способность выйти за рамки причинных связей, установленных разработчиками, и возможность «осознавать» собственные действия* [McCarthy, 1995-1].

Имеются две трактовки теста Тьюринга с позиции «здравого смысла». Первая, широкая трактовка, превалирующая в дискуссиях в 1950 – 1975 гг., касается тезиса Тьюринга о возможности замены вопроса «Может ли машина мыслить?» на вопрос «Может ли машина играть в игру в имитацию интеллектуального поведения?». Если постановка такой игры бессмысленна, то нечего утверждать о выявлении посредством ТТ каких-либо ЭЭ-приспособлений в компьютерной системе. Вторая трактовка, более узкая, предполагает логическую эквива-

лентность ответов на данные вопросы и направлена на спецификацию теста, выявляющего «проблески» «здорового смысла» в χ -системе. Сегодня, по всей видимости, в связи с впечатляющими успехами ИИ, термин «здоровый смысл» используется лишь в контексте узкой трактовки, впервые предложенной в оригинальном ТТ.

Эволюционно-эпистемологические идеи в оригинальном тесте Тьюринга

В тьюринговой идее игры в имитацию интеллекта, особенно в части, касающейся обоснования инженерного проекта компьютера, чётко выявляются две отмеченные нами эпистемологические ветви: УЭ- и ЭЭ-парадигма. Первая инспирирована проектом универсальной цифровой вычислительной машины (УЦВМ), которая, напомним, является развитием идеи машины Тьюринга – концептуального инженерного проекта, включающего в себя детерминированное управляющее устройство, переходящее из одного состояния в другое, которое перемещается по ленте, считывая с неё и записывая на неё символы. Программа функционирования машины представлена в автоматной таблице. УЦВМ задаёт формальное определение квазиалгоритма благодаря механизму случайных переходов по автоматной таблице и наличию книги правил, на каждой странице которой (на ленте) представлен частный алгоритм. Также напомним, что, помимо стандартных алгоритмических свойств, УЦВМ обладает индетерминированностью, сингулярностью, нерезультативностью, непрерывностью.

А.Тьюринг совершенно справедливо посчитал, что мощности УЦВМ не достаточно для реализации механизма прироста знаний, и прорисовал контуры ЭЭ-парадигмы в идее построения обучаемых машин на базе нейрокомпьютера. Такое решение было принято против аргумента А. Лавлейс: «Аналитическая машина (универсальная ЦВМ – А.А.) не претендует ни на что оригинальное. Что бы машина ни сделала, нам известен способ, как это сделано» [Turing, 1950, P. 450]. То есть машины не способны к самостоятельному творчеству, так как не могут создать ничего нового, они способны только следовать программным инструкциям. Всё «творческое» предопределено разработчиком.

Таким образом, ЭЭ-проблематика в оригинальной версии ТТ инициирует общую проблему творчества в ИИ и требует специального освещения (см. подразд. 3.2). Отметим лишь основные пункты данной

проблематики. А.М. Тьюринг предлагает два варианта ответа на возражение А. Лавлейс – «репрезентативный» и «коннекционистский» [Алексеев, 2011-8].

В рамках *репрезентативного опровержения* А.М. Тьюринг апеллирует к проекту УЦВМ, что приводит лишь к схоластическим изысканиями. Например, он утверждает: «Под солнцем ничего нового не происходит. Всё, что мы считаем оригинальным, на самом деле представляется лишь результатом обучения или следствиями общеизвестных принципов» [Turing, 1950, P. 450]. Человеческая способность к творчеству совершенно непонятна. Удивление нетривиальностью результатов работы программы из-за сбоя машины сродни удивлению творческим гением. (Репрезентативная аргументация будет подробно рассмотрена в подразд. 3.2.)

Очевидна логическая некорректность аргументации, как правило, из-за подмены тезисов. Так же опровержение неубедительно из-за объективных факторов – символьными средствами УЦВМ трудно и, по всей видимости, невозможно адекватно имитировать креативные феномены. Логико-атомистический каркас репрезентативной парадигмы не обеспечивает свободы развития сущностей и связей моделируемой предметной области, ведь любая достаточно сложная система формализованных знаний продуцирует результаты в контексте известного (здесь срабатывает аргумент Гёделя). Комбинация экспертных знаний не ведёт к росту знаний. Доводы в пользу креативного лингвокомпьютинга, подкреплённые проектом УЦВМ, не убедительны.

Иное дело – *коннекционистское опровержение*. Предлагается концептуальный проект машины, снабженной нейросетевой программой обучения. Концепция нейрокомпьютинга коренным образом меняет исходную идею ТТ – от имитации интеллекта А. Тьюринг отказывается и настаивает на моделировании ЭЭ-механизмов развития человеческого интеллекта. Интерпретируя не совсем чёткие коннекционистские взгляды А. Тьюринга, предложим следующий вариант тьюринговой методики разработки обучаемой машины: 1) построение исходной системы знаний и интеллектуальных способностей, которыми обладает взрослый человек; 2) изучение способа роста знаний в ходе обучения; 3) изучение опыта, который не может быть результатом обучения; 4) очистка от содержания «знаний» и формирование структурно-

процедурального скелета интеллектуальных способностей; 5) представление «чистых» знаний как модели интеллекта ребёнка – *машины-ребёнка*; 6) обучение машины-ребёнка и формирование *машины-взрослого*; 7) сравнение «знаний» и «способностей» машины-взрослого со знаниями и способностями взрослого человека. Важно то, что учитель, как правило, не осведомлён о внутренних процессах машины-взрослого. Тем не менее, в некоторой степени он может предсказывать поведение «ученика». Именно в этом А. Тьюринг усматривает опровержение аргумента Лавлейс, одна из версий которого: «машина может делать только то, что мы знаем, как приказывать ей это сделать» [Turing, 1969]. Возможно, это так, однако в случае с обучаемой машиной мы не «знаем» наверняка о её поведении, но лишь прогнозируем его и проверяем гипотезы.

А. Тьюринг полагает, что машину-ребёнка создать достаточно просто. «Интеллект ребёнка подобен чистому блокноту, купленному в магазине канцелярских товаров». Поэтому достаточно разработать механизм развития «знаний» машины-ребёнка и иметь большое количество «чистых листов» (для полного представления предметной области). Конечно, это неверно. Сегодня стало общим положением утверждение о том, что сложность моделирования процесса развития знаний ребёнка на несколько порядков выше сложности модели мышления взрослого человека. А. Тьюринг в данном вопросе оптимист и предлагает несколько интересных методов обучения машины-ребёнка, например, «метод кнута и пряника» [Ibid, P. 17 – 23].

Машина-ребёнок реализуется нейронной сетью. По поводу устройства машины-взрослого остаётся только догадываться. Собственно, и концепция нейронной сети у А. Тьюринга не убедительна. Если её более чётко рассмотреть, то никакого нейрокомпьютера мы не увидим, но лишь параллельную УЦВМ. Чистые листы в блокноте машины-ребёнка (книга правил УЦВМ), точнее, стопка перфокарт (пусть это будут «слои» нейронной сети) и механизм параллельного доступа к ним (присваивающий «веса» «нейронам» из каждого «слоя») претендуют лишь на формальное определение параллельного компьютеринга, но никоим образом не нейрокомпьютеринга. Несмотря на неясность конструкции, А. Тьюринг был убеждён: мыслящую машину следует строить как человеческую модель обучения [Ibid, P.14]. Именно к обучаемым компьютерам относится заключительный и часто цитируемый вывод

А. Тьюринга относительно роли и места интеллектуальных машин в обществе как соперников и помощников человеку во всех сферах деятельности.

Несомненно, коннекционистское опровержение аргумента Лавлейс намного убедительнее репрезентативного, особенно если аргументировать его более правдоподобным проектом нейросетевой машины, фундаментальные принципы функционирования которой не стоит связывать с линией «Бэббидж – Тьюринг», но с внедрением машины Корсакова в концептуальный аппарат перспективного компьютеринга (см. Заключение). Пусть машина Тьюринга останется главной для лингвокомпьютеринговых принципов функционирования интеллектуальных машин и для УЭ-возможностей роста их знаний. Эти возможности весьма ограничены, что выявляется при изучении теста Мура, в котором ТТ специально предназначался для развития компьютерных «знаний» о компьютерных «знаниях».

Тест Мура: ущербность эволюционно-эпистемологической парадигмы традиционного искусственного интеллекта

Дж. Мур в статье «Анализ теста Тьюринга» (1976 г.) в тьюринговой игре в имитацию усматривает в большей мере практический, нежели теоретический ориентир. Главный тезис: «Тест Тьюринга только тогда значим для компьютерной технологии, когда он проинтерпретирован в индуктивном формате» [Moog, 1976]. Тест Тьюринга – это базис индуктивного доказательства гипотезы в пользу компьютерного мышления. Надо начать работу по систематическому сбору и обобщению фактов прохождения частных аспектов ТТ. Как в процессе автоматического дедуктивного доказательства истинности/ложности теорем получается «побочный продукт» – значение терминов рассуждения, так в ходе индуктивного доказательства компьютерная система не только утверждает ту или иную интеллектуальную способность, но и самостоятельно расширяет собственные «знания». Дж. Мур приводит ЭЭ-соображения в стиле бадеёйной теории обучения: 1) ТТ идентифицирует, систематизирует и обобщает факты реализации интеллектуальных функций у машин; 2) ТТ поощряет жёсткий отбор и отбраковывает непригодные факты; 3) положительный и отрицательный опыт сохраняется в памяти компьютерной системы, которая становится всё более «интеллектуальной» и требовательной к новым прецедентам прохождения ТТ. К сожалению, Дж. Мур

не предложил конструктивного проекта системы, способной к самособиранию фактов интеллектуальности и самоизменяющейся в соответствии с приростом «знаний». В частности, это объясняется теоретической неразрешимостью проблем индукции и фреймов. Например, что общего, между игрой в шахматы и принятием решения? Очевидно, одно другому способствует, но как? Поэтому тест Мура следует считать лишь заявкой на ЭЭ-проект компьютерной системы.

Тест Колби: универсально-эпистемологическая парадигма ИИ и параноидальные системы

Мощным критическим зарядом относительно УЭ-парадигмы развития «знаний» экспертных систем обладает параноидальный ТТ, предложенный в исследованиях К. Колби и его коллег. В статье 1971 г. «Искусственная паранойя» [Colby, Hilf, Weber, 1971] описывается компьютерная программа PARRY, имитирующая поведение пациентов психиатрической больницы. Тест Колби – это не чистый мысленный пример, но реальная инженерная конструкция, которая наряду с программой ELIZA является одним из первых примеров удачной практической реализации ТТ. Результаты тестирования ошеломили многих: почти в половине случаев психиатры не могли отличить компьютер от больных. Ответы сопровождались словосочетаниями, окрашивающими речь эмоциональными состояниями страха, гнева и недоверия, что во многом способствовало имитации параноидального бреда.

В следующей статье «Роль неразличимости в тьюринго-подобных тестах для компьютерного моделирования параноидальных процессов» [Colby, Hilf, Weber, Kraemer, 1972] исследователи акцентировали на теоретических сравнениях своего теста и теста Тьюринга. И оригинальный и параноидальный тест – это инструмент, обосновывающий правомочность имитации различных форм человеческой деятельности, будь она рациональна или нет. Например, параноидальный тест легко пройдут те, кто оспаривает ТТ, доказывая, что машина не может мыслить, при этом замыкаясь, как параноики, на некотором частном определении «сущности» интеллекта.

Моделирование паранойи оказалась дорогим занятием: PARRY – очень «увесистая» программа, словари включали 4 500 слов, справочники – 700 статей. Программа синтаксического анализатора вопросов и грамматического синтезатора ответов превышала 200 000 инструкций.

«Зачем идти на такие затраты и моделировать параноидальное поведение? Какова была цель теоретических и практических изысканий К. Колби? Не прошло ли всё это впустую и не выкинуты ли деньги на ветер?» – задаёт вопросы Д. Деннетт [Dennett, 1984] и предлагает очень интересное сравнение параноидальной программы PARRY с экспертными системами, стоимость которых на порядки выше, но которые способны пройти лишь несколько усовершенствованный параноидальный тест – из этого теста изъяты выражения параноидальных эмоций.

Тест Деннетта: оперативное тестирование эволюционно-эпистемологических механизмов экспертных систем

В работе «Эра интеллектуальных машин. Могут ли машины мыслить?» [Ibid] Д. Деннетт с позиции идей теста Колби анализирует экспертные системы, которые стали сегодня классическими прототипами и изучаются во всевозможных вузовских курсах искусственного интеллекта. PARRY – это компьютерная модель теории психологического феномена – паранойи – и на практике не применима. Тем не менее, экспертные системы, возникшие в «инженерии знаний» и пользующиеся большим практическим спросом, по способу прохождения ТТ аналогичны тесту на паранойю. Несомненно, подчёркивает Д. Деннетт, экспертные системы – это превосходные консультанты-специалисты, которые способны поставить медицинский диагноз, изучить геологические данные, проанализировать результаты научных экспериментов, отреагировать на новости. Например, интерес представляет новостная экспертная система CYRUS (САЙРУС), которая моделировала «знания» Сайруса Вэнса (Cyrus Vance) – секретаря в администрации американского президента Картера. Цель проекта CYRUS была амбициозной – стать «чистой системой искусственного интеллекта», искусственным механизмом формирования и развития «знаний», аналогичного человеческому когнитивному механизму формирования нарративного сознания, т.е. сознания личности, основанного на биографических воспоминаниях. CYRUS обновлялась каждый день из блоков новостей американских информационных агентств. Как только в новостях появлялось ключевое слово «Cyrus Vance», то ассоциированная с ним информация оперативно записывалась в базу данных CYRUSa. По мере накопления данных возрастала способность CYRUSa отвечать на всё большее и

большее количество вопросов. К CYRUS обращались во втором лице, как при личном разговоре с человеком Сайрусом Вэнсом.

Пример диалога: «Деннетт (Д): В прошлый раз, когда вы ездили в Саудовскую Аравию, где вы останавливались? Сайрус (С): Во дворце в Саудовской Аравии 23 сентября 1978 г. (Д): Вы ходили смотреть там достопримечательности? (С): Да, на месторождение нефти в Хоране 23 сентября 1978 г. (Д): Встречали ли Вы мистера Бегина? (С): Да, последний раз на торжественном обеде в Израиле в январе 1980 г.»

Кажется, что система CYRUS способна правильно отвечать на все вопросы, которые только можно придумать и корректно сформулировать. Однако её легко сбить, если задать глупый вопрос.

Пример теста на «здравый смысл». Д. Деннетт задаёт CYRUS вопрос: «Вы когда-нибудь встречали женщину, руководящую государством?», проверяя, «знает» ли CYRUS, что Индира Ганди и Маргарет Тэтчер были женщинами. CYRUS не может ответить ни да, ни нет. Экспертная система оказалась в тупике. Но ведь вопрос был достаточно простой, неужели в экспертной системе не предусмотрены подобные вопросы, которые требуют примитивных рассуждений? Причина несуразицы очевидна. Предлагается взглянуть во «внутренность» системы CYRUS. Обнаруживаются «скелетные» описания тысяч слов, минимальные настолько, насколько решили когнитологи (Д. Деннетт под «скелетными описаниями» понимает широко известные среди специалистов контекстно-зависимые скрипты Р. Шенка [Шэнк, 1980]). Например, «адвокат» описывается как синоним слова «юрист». Однако система «не знает», что «адвокаты» – это взрослые люди; в сфере закона они выполняют различные функции; являются выпускниками университетов; платят им больше, чем горничным; знают, как завязывать шнурки; вряд ли будут искать компанию среди дровосеков и т.д. Это – банальные, но очевидные факты об адвокатах. В CYRUS эти знания отсутствуют. Для представления «здорового смысла» требуется слишком много понятий, характеризующих жизненный опыт человека, и, как только встречается серьёзное отклонение от вложенных в систему правил, – таких отклонений бесконечно много, система перестает работать и требует доработки для ответа на элементарные вопросы с позиции «здорового смысла».

Вердикт Д. Деннетта уничижителен – современные классические интеллектуальные системы способны пройти лишь параноидальный тест Тьюринга. Все они – «деревни Потёмкина»: у них интересная внешняя оболочка – привлекательный интерфейс, как фильм в кинотеатре, однако внутри ничего нет. Экспертные системы, подобные

CYRUS, интересны лишь в теоретическом плане – как *не* надо строить интеллектуальные системы.

Тест Маккарти: долгосрочное тестирование эволюционно-эпистемологических механизмов экспертных систем

Идею компьютерных программ, имитирующих «здравый смысл», Дж. Маккарти прорабатывает более полувека, начиная с 1958 г., когда впервые обнаружил её при формализации контекста логических рассуждений [McCarthy, 1969]. Дж. Маккарти солидарен с Д. Деннеттом в том, что показатели функционирования экспертных систем в отдельных областях знаний впечатляют. Однако не следует такие системы относить к «интеллектуальным». Это – обычные программы, и уровень их «интеллекта» не повышается за счет применения высокоуровневых языков программирования или использования нечётких моделей. В основе теста Маккарти лежит работа «Экспертные системы нуждаются в здравом смысле» [McCarthy, 1984], опубликованная почти одновременно с представленной выше работой Д. Деннетта. Тест отличается от деннеттовской «быстрой проверки»: длина теста может быть значительной, анализ – скрупулёзным. Если для Д. Деннетта «здравый смысл» характеризует банальные, повседневные вещи и положения, то для Дж. Маккарти это – показатель богатого профессионального опыта и глубоких теоретических знаний специалиста.

Предметом критики послужила экспертная система MYCIN. Программа консультирует врачей по вопросам лечения бактериальных инфекций крови и менингита. Как правило, MYCIN показывает гораздо более высокие результаты по сравнению с заключениями студентов-медиков, интернов и даже специалистов по бактериальным заболеваниям. Работает достаточно надежно и, на первый взгляд, для её эффективного функционирования излишни изыскания по поводу компьютерной реализации понятия «здравый смысл». Это не так. Её широкое применение застопорилось из-за следующих причин: 1) требуется регулярное пополнение базы «знаний» не только наличными, но и неизвестными данными; 2) система не способна самостоятельно определять собственные ограничения.

Пример теста на «здравый смысл». В систему вводят данные о том, что у такого-то пациента обнаружено поражение кишечника холерным вибрионом. MYCIN бодро выдаёт рекомендации о проведении

двухнедельного курса лечения тетрациклином. Если слепо довериться рекомендациям, то бактерии будут уничтожены, однако вероятнее, вначале пациент скончается от холеры. Летального исхода можно избежать, если врач-пользователь MYCIN знает, что данный случай диареи требует интенсивного лечения, и предпримет дополнительные меры, не учтённые в экспертной системе, для выяснения соответствующих методов и препаратов. Но это будет «здравый смысл» врача-человека, обладающего большим опытом лечения подобного рода болезней. Отсюда тезис: «В этой экспертной системе нет ни толику здравого смысла».

Для прохождения теста следует ответить на большое количество разнородных вопросов, сводимых к двум классам: 1) «что» такое «здравый смысл»?; 2) «как» моделировать рассуждения на основе имеющихся знаний о «здравом смысле»? Ответы на эти вопросы сводятся к следующему: 1) «здравый смысл» тесно связан с конкретикой реальных событий и действий, поэтому система должна уметь самостоятельно прогнозировать последствия собственных действий; 2) «здравый смысл» – это рефлексия над конкретными действиями; 3) для разумного поведения в сложных условиях «здравого смысла» недостаточно, требуется последующая рефлексия, умение представлять и использовать *знания о знаниях*; 4) необходимо учитывать информацию о личностях – их знаниях, убеждениях, целях, симпатиях и антипатиях, намерениях и возможностях и пр.; 5) «здравый смысл» опирается на способы формализации «наивной» физики, которые объёмнее научной физики. Например, всем известно что произойдёт, если бросить стакан воды с метровой высоты на пол: стакан разобьётся; на это уйдут доли секунды; брызги воды не разлетятся дальше нескольких метров и т.п. Наивные знания не включают формулы падения тела и уравнения течения жидкости, исходных данных недостаточно, и поэтому на решение уйдёт много времени. Если «знания» робота – это теории механики, то он не успеет сойти с дороги, когда навстречу ему мчится автомобиль; 6) «здравый смысл» составляют естественные понятия (виды), которые формируются в обыденной практике. Например, понятие «шум» – это и лай собаки и падение капли воды. Проблема представления подобных фактов *универсальным способом* не разрешима; 7) имеется также «проблема фреймов» – как из частных спецификаций получить целостное описание предметной области.

Несмотря на принципиальные теоретические трудности моделирования «знаний» о «здравом смысле», Дж. Маккарти предлагает ряд частных конструктивных решений, например, логику формализации контекста («смысл» как контекст «значений» языковых выражений) [McCarthy, 1982], язык программирования экспертных систем, основанный на теории речевых актов [McCarthy, 1986], и пр.

Тест Деннетта и тест Маккарти дополняют друг друга в смысле разных трактовок понятия «здравый смысл». Но как бы ни различались подходы, напрашивается такой вывод – экспертная система, реализующая тест на здравый смысл, должна быть искусственной личностью и жить жизнью людей. Конечно, это – фантастика дурного вкуса, особенно если учесть отсутствие каких-либо достижений в программировании ЭЭ-механизмов ИИ-систем.

Эволюционно-эпистемологические механизмы динамических экспертных систем реального времени

В современной методологии экспертных систем наблюдается более удручающее положение в области развития «знаний», нежели чем во время классических систем, на которые обратили внимание Д. Деннетт и Дж. Маккарти. Начиная с середины 1990-х годов, в сложных информационных технологиях стали доминировать *динамические экспертные системы реального времени* – G2, RT Works, TDC Expert и др. «Именно они позволили искусственному интеллекту перейти от игр и головоломок к массовому использованию при решении практически значимых задач» – считает Э.В. Попов, видный отечественный исследователь ИИ [Попов, 1995]. Система дорогая, и для некоторых предметных областей ее цена достигает сотен миллионов долларов. Такая стоимость оправдывается не столько ноу-хау программ, сколько армией разработчиков, включённых в технологию её непрерывного сопровождения и интерактивно связанных с пользователями. Система способна быстро реагировать на динамику «знаний», однако всякий раз, когда изменение форм и способов представления «знаний» превышает учтённые границы, требуется вмешательство программиста. Имеется много технических несуразиц, которые следует критически изучить. Например, странно то, что запрещается пользоваться классическими языками представления «знаний», такими, как LISP, первый вариант которого разработал Дж. Маккарти более полувека назад.

Версия абсолютной динамической экспертной системы в точности воспроизводит концепцию машины Блока – в ней максимально задействован потенциал УЭ-парадигмы развития «знаний». Совершенная система объединяет разработчиков всех стран в рамках глобального виртуального интернет-сообщества, адресно доставлять им запросы пользователей, на которые они в режиме реального времени способны мгновенно ответить, задействовав всю мыслимую мощь информационно-коммуникационных технологий. Несомненно, система проходит тест Тьюринга, в том числе и тест Деннетта и тест Маккарти – кто-то из человечества обязательно найдется, чтобы ответить на вопросы, заданные с позиции «здравого смысла».

Но является ли она интеллектуальной искусственной системой? Совершенно очевидно – нет. Здесь нет «ни грамма» искусственного интеллекта, но лишь интеллект разработчиков. Более того, так как знания разработчика фундированы некоторым личностным смыслом, частным для общечеловеческого формата, то абсолютный вариант динамической экспертной системы – это *глобальная параноидальная система*. Кстати, этому прототипу соответствует Интернет, так как стратегия развития поисковых фирм типа Google и Yandex – интеллектуализация веб-навигации.

Крайне пренебрежительно сторонники динамических систем отзываются по поводу *статических экспертных систем*, названных так из-за отсутствия связи с разработчиком. На мой взгляд, в искусственном интеллекте всё должно быть с точностью наоборот. ИИ-система призвана работать автономно. Независимо от разработчика она должна извлекать, приобретать, репрезентировать, хранить, обрабатывать, интерпретировать, выдавать, передавать и – самое главное – самостоятельно развивать «знания».

Причина мощного коммерческого успеха «динамических» систем видится в обычном PR-трюке – на брэнд интеллектуальности «ключают» заказчики и удобно маскировать машинную «глупость» такими, как «экспертные системы реального времени – основное направление искусственного интеллекта!» [Попов, 1995]. Конечно, в этих системах есть многое: открытость, масштабируемость, объектно-ориентированное программирование, CASE-технологии, клиент-серверные технологии, рассуждения на базе прецедентов и на базе ими-

тационных моделей, параллельное выполнение в реальном времени, когнитивная графика, технологии «мягких» вычислений, структурированный естественный язык представления «знаний», модели внешней и внутренней среды и многое другое. Тем не менее, мы имеем полное право утверждать, перефразировав Дж. Маккарти:

**В экспертных динамических системах реального времени нет
ни толики искусственного интеллекта**

Искусственный интеллект вне эволюционно-эпистемологической парадигмы

Чтобы хотя бы в заключение данного материала прозвучали не только критические, разрушающие глобальную индустрию «интеллектуальных динамических экспертных систем», но и конструктивные идеи, вернёмся к проекту компьютера, способного реализовать комплексный Тест Тьюринга. Для данного проекта УЦВМ оказалась слабой, а нейронная сеть Тьюринга – всего лишь усовершенствованной параллельной УЦВМ. Предлагаем обратить внимание на крайне перспективный, хотя более старый проект, предложенный за 120 лет до статьи Тьюринга, в 1832 г., российским инженером С.Н. Корсаковым. Машина Корсакова, по сути, является протонейрокомпьютером, и собственно с неё предлагается отсчитывать идею нейрокомпьютинга в мировом масштабе [Алексеев, 2011-7]. Метафора такова: стопка перфокарт прокалывается иглами, и глубина проникновения игл характеризует связь между ассоциируемыми понятиями, обозначенными на перфокартах, т.е. «вес коннекции» задаётся усилием прокалывания, но не символом, который считывается с перфокарты или записывается для варианта параллельной УЦВМ. Имеется ряд интересных исследований машины Корсакова, которые меняют фундаментальные принципы современного компьютеринга посредством комплексирования машины Тьюринга и машины Корсакова [Алексеев, 2011-3; 2012-1]. Например, формальное тьюринговое определение алгоритма работы математика дополняется «алгоритмом» деятельности художника, где интенсивность прокалывания иглами перфокарт задаётся эмоциональным восприятием предметной области, которая репрезентирована «явными знаниями» на текущей перфокарте, а также «неявными» и «фоновыми» «знаниями», обозначенными на других, ниже и выше лежащих перфокартах. Несомненно,

мненно, это требует специального доказательства и формальной экспликации. Но даже с первого взгляда видно, что лингво-нейро-машина Тьюринга-Корсакова меняет теоретико-алгоритмические основания компьютеринга, включая принципы реализации эволюционно-эпистемологических теорий.

Критический результат данной статьи следующий. Если на заре развития ИИ эволюционно-эпистемологическая парадигма развития «знаний» вызывала острый экспериментальный интерес, то сегодня страсти поутихли под массивным напором универсально-эпистемологических технологий. Несомненна сиюминутная результативность УЭ-парадигмы, поэтому финансовые вложения в неё огромны. Однако в последующем УЭ-парадигма заводит в тупик, и исследования ИИ сводятся на нет. Фундаментальная эпистемологическая задача ИИ – создание механизмов саморазвития «знаний» компьютерных систем – остаётся в стороне от серьезных крупных междисциплинарных исследований. Непосредственной отдачи и рыночной выгоды от её решения ждать не стоит. Поэтому внедрение идей ЭЭ в ИИ и, собственно, развитие истинного, настоящего, неинновационного ИИ осуществляется спорадически энтузиастами. Хотя, как мы показали, достаточно простых рассуждений на тему теста Тьюринга, чтобы разрушить мнимый образ «интеллектуальности» динамических экспертных систем и показать, что сторонники УЭ-парадигмы занимаются чем-то другим, но не искусственным интеллектом. Так как комплексный ТТ прорисовывает фундаментальные, концептуальные, мыслимые очертания ИИ, то вполне закономерен вывод:

ИИ без ЭЭ не мыслим

5.2. Критика методологии: комплексный тест Тьюринга против комплексного проекта искусственной личности

В последние пятнадцать лет в философии искусственного интеллекта обозначилась область междисциплинарных исследований, получившая название «Проект искусственной личности». Возникновение проекта во многом объяснимо факторами общественной жизни – сегодня сложные интеллектуальные информационно-технические системы невозможно рассматривать с сугубо технической точки зрения, они приобретают социокультурный, человекомерный статус. Используемые в настоящее время и перспективные информационные технологии рассматриваются в контексте социальных ценностей, мировоззренческих ориентиров, морально-правовых норм, эстетических канонов и иных составляющих духовной сферы. Здесь просматривается требование, описанное в отечественной философской науке как необходимость постнеклассической переориентации методологии изучения, построения и развития сложных систем [Степин, 1991]. Однако в большей мере возникновение проекта искусственной личности обусловлено внутриметодологическими факторами, которые отражают достижения искусственного интеллекта. От моделирования чисто интеллектуальной деятельности осуществлён системологически закономерный переход к моделированию сознания. Когнитивно-компьютерные модели сознания стали рассматриваться в качестве концептуального и реализационного базиса «персонализации» интеллектуальных информационных систем [Алексеев, 1998].

Дефиниция искусственной личности

Искусственная личность (ИЛ) – это когнитивно-компьютерная система, удовлетворяющая, по крайней мере, одному из двух критериев: 1) она обладает подобием человеческой субъективной реальности – так называемым «квасисознанием» (в ряде работ употребляется термин «псевдосознание»); 2) она функционально, поведенчески и/или физически *неотличима* от человеческой личности.

Первый критерий ИЛ поддерживается, в основном, сторонниками сильного ИИ. Они считают, что для реализации сложной функциональной (само-)организации компьютерной системы и осуществления ею адаптивного поведения в условиях динамики внешней и внутренней

среды необходимо выделение в архитектуре системы специального блока «квазисознания», системологическая роль которого подобна роли сознания в жизни сознательного существа. Этот блок должен осуществлять проективные, интроспективные, квалиативные и прочие отношения с внешней и внутренней средой системы. Критерий правдоподобен в контексте парадигмы компьютеризации. Эта разновидность функционализма по-иному называется «функционализмом машины Тьюринга», «функционализмом искусственного интеллекта», «вычислительной теорией сознания». Правдоподобие критерия, однако, шатается из-за многочисленных шагов редукции: так, персональное редуцируется к ментальному, ментальное – к интеллектуальному, интеллектуальное – к структурно-функциональной архитектуре компьютера. Формальные зависимости между когнитивными и компьютерными компонентами задаются посредством вычислительных элементов, операций и функций, а каузальные зависимости находят наиболее общее выражение в реализационных аспектах машины Тьюринга. Вычисление понимается расширенно – репрезентативный и коннекционистский подходы объединяются. Полученные при этом интегральные репрезентативно-коннекционистские кодовые структуры операционализируются средствами квазиалгоритмической обработки. В результате многошаговой редукции получаем компьютеризационистскую формулировку проекта ИЛ: *личностное (персональное) – это квазиалгоритмическое вычисление.*

Во втором критерии ИЛ показателем «неотличимости» естественной и искусственной личностей выступает способность системы пройти *полный тест Тьюринга* (ТТ) [Алексеев, 2006-1] В этом тесте, помимо вербально-коммуникативной, перцептивно-моторной, анатомо-физиологической и даже (в некоторых разновидностях ТТ) микрофизической неотличимости искусственной и естественной систем, всесторонне учитываются поддающиеся внешнему наблюдению показатели духовной сферы деятельности личности – персонологические параметры. К ним относятся «смысл», «свобода», «любовь», «ответственность», «право», «творчество», «красота» и др. Правдоподобие второго критерия также, как и первого, проблематично. Здесь особенно не ясна связь между внутренним и внешним миром человека. С такой проблемой почти век назад столкнулся бихевиоризм. Несмотря на очевидные заслуги в устранении метафизики ментальных сущностей широко принимается обстоятельная критика бихевиорального метода оценки сознания исхо-

дя из наблюдений за действительным поведением и из предположений возможного поведения.

В таких условиях – условиях проблематичности критериев оценки интеллектуальных систем как систем искусственной личности – развиваются современные проекты.

Многообразие проектов искусственной личности

Можно показать, что история философии искусственного интеллекта – это история обсуждения возможностей компьютерной реализации персонологических параметров. Достаточно вспомнить полемический стандарт А. Тьюринга по поводу построения мыслящих машин – совокупность аргументов и контраргументов в решении «основного вопроса» философии искусственного интеллекта: «Может ли машина мыслить?». Эту совокупность мы называли «стандартом» по причине придания формы и отчасти содержания современным дискуссиям в области философии искусственного интеллекта. В «стандарт» вошли аргументы и контраргументы явно не инженерно-технического характера: теологический, антисциентистский, креационистский, «от первого лица», «от другого сознания» и даже экстрасенсорный. Целевое – персонологическое – назначение проектов вычислительных машин «идёт от века». Достаточно вспомнить о машине Р. Луллия (1235 – 1315). Здесь мы видим не только обсуждение вопросов логики оперирования понятиями, но и религиозно-моральное назначение проекта «Арс магна» [Реале Антисери, 1997, С. 206 – 207]. Так же и для Лейбница при построении вычислительной машины второстепенными были вопросы логико-математического плана, а первостепенными – юридическо-правовые вопросы.

С начала 1990-х годов выделяются следующие крупные проекты ИЛ [Алексеев, 2005-10]: 1) OSCAR Дж. Поллока, сформулированный в рамках «универсальной теории рациональности» и её приложениями для построения искусственных рациональных агентов («артилектов») [Серёдкина, 2007]; 2) проект «человекоподобных агентов» А. Сломана, призванный реализовать широкий спектр персонологических параметров, например «любовь», «свобода» [Sloman, 1978; Sloman, Croucher, 1981-2; Sloman, 1996; 1998; 2000]; 3) проект гуманоидных роботов КОГ, в котором Д. Деннетт усматривает апробацию собственной теории множественных набросков, где персональное возникает из сложного сочетания бесчис-

ленных серий нарративов, а личность и социум – это субстанциональные системы бесчисленных роботов, в которых «ментальное» представляется компонентой функциональной самоорганизации [Деннетт, 2004]. В этих проектах предлагаются концептуальные, логико-математические, программные решения. Нам интересен методологический уровень осмысления путей реализации проектов ИЛ.

В связи с этим и в соответствии с введёнными критериями ИЛ показательными для проектов ИМ представляются два подхода [Алексеев, 2008-4; 5]: 1) Д. Лената, автора широко известной программы «Автоматический математик» (1976 г.), который считает, что следует подражать психологическим, социокультурным, лингвистическим, интеллектуальным и другим особенностям личности, а физическое подобие системы – второстепенный, несущественный фактор; 2) подход Р. Брукса, ученика Д. Лената, автора вышеупомянутого робота КОГ и знаменитой версии *Kismet*, имитирующей мимику человеческого лица. Р. Брукс полагает, что физическая антропоморфность робота – первичное и необходимое качество его персонологического подобия, ведь все социокультурные понятия, на основании которых будет функционировать робот и которые он будет вырабатывать в ходе адаптации к внешней и внутренней среде функционирования, предопределены антропологическими особенностями человека. Например, согласно принципу Джонсона и Лакоффа, «безногому» роботу трудно будет усвоить выражение «Перевернуть с головы на ноги».

Подход Д. Лената и Р. Брукса можно обозначить соответственно как *экспертный* и *робототехнический* подходы к построению ИЛ.

В англо-американской философии ИИ в основном преобладает робототехнический подход, согласно которому ИЛ – это обладающий квазисознанием автономный робот. Экспертная методология не признаётся такими авторитетными зарубежными философами ИИ, как Дж. Маккарти А. Сломан. В отечественной же науке, напротив, сложился экспертный подход: ИЛ рассматривается в социально-эпистемологическом контексте междисциплинарных взаимодействий специалистов, в ходе которых формируются данные и знания социокультурного содержания. Рассмотрим более подробно эти подходы.

Искусственная личность – робототехническая система

Искусственная личность – это робот, наделённый квазисознанием – т.е. совокупностью, по крайней мере, некоторых персонологических способностей и качеств человеческой личности. Здесь показательна дискуссия по поводу работы С. Брингсйорда – «Чем могут и не могут быть роботы» (1992, 1994 гг.) [Bringsjord, 1994], продолженная на страницах сайта <http://psycprints.ecs.soton.ac.uk> (рубрика «Сознание робота»). В этой работе позитивные утверждения относительно проекта ИЛ сопровождаются обстоятельной критикой. Основной девиз работы звучит так: «В будущем робот будет делать всё то, что делаем мы, но не будет одним из нас» – т.е. не будет сознательным.

С. Брингсйорд доказывает следующее: а) ИИ будет производить машины, обладающие способностями проходить всё более и более сильные версии ТТ; б) проект ИЛ по созданию машины–личности будет неминуемо проваливаться. В защиту (а) предлагаются компьютерные программы генерации художественных и философских текстов и индуктивный вывод при построении версий ТТ: тьюринговая игра в имитацию → наблюдение внешнего вида игроков → изучение их сенсомоторного поведения → сканирование мозга и пр. Автор уверен, что ИИ будет мало-помалу проходить данную последовательность. В основе доказательства (б) – несостоятельности проекта ИЛ – лежит *modus tollens* из суждений:

- (1) «Проект ИЛ» → «Личность – это автомат»;
- (2) ¬ «Личность – это автомат»; ⇒
- (3) ¬ «Проект ИЛ».

Посылка (2) интуитивно апеллирует к наличию у человека и отсутствию у автомата ряда машинно невозпроизводимых персонологических параметров F, к которым относится свободная воля, способность к интроспекции, внутренний проективный опыт «каково быть» («what it's like to be») и пр. К F относятся и способность к компьютерному воспроизводству творческих способностей. Эту способность автор подробно анализирует в ряде предыдущих работ (в тесте Лавлейс (см. подразд. 2.2) и делает вывод «Компьютер творить не может!» (ранее в тесте Лавлейс отмечался подход С. Брингсйорда). Общее заключение:

- (4) «Личность обладает F»;
- (5) «Автомат не обладает F»; ⇒

(6) «Личность не может быть автоматом».

С. Брингсйорд достаточно подробно обосновывает невычислимость F, апеллируя к аргументу Гёделя, «Китайской комнате» Дж. Серля, аргументу произвольной множественной реализации Н. Блока и др. Итог таков: *«Роботы будут многое делать, однако они не будут личностями»*.

В полемике по поводу работы С. Брингсйорда наблюдается четыре направления критики и аргументации относительно реализации проекта искусственной личности¹:

1. Проект ИЛ *абсурден*, так как: 1) метафизические понятия «личность», «свобода воли», «интроспекция» не могут быть предметом эмпирического анализа; 2) нечёткое определение понятия личности, образная иррациональность проекта ИЛ – всё это не заслуживают научного внимания; 3) персонологические параметры логически невыразимы – попытки их логической экспликации непременно сопровождаются техническими ошибками.

2. Проект ИЛ *неточен*, так как: 1) апелляция к интуитивным аргументам разрушает логическую строгость дедуктивных аргументов; 2) нецелесообразно применять дедуктивные умозаключения в качестве метода аргументации (дедукция – не метод, а схема суждений); 3) необходимо чётко формулировать конкретную разновидность функционализма, который лежит в основании проекта ИЛ – нет функционализма

¹ См. дискуссию на сайте <http://psycprints.ecs.soton.ac.uk/archive/00000418/> следующих авторов: Barresi, John (1995) Building Persons: Some Rules for the Game; Bringsjord, Selmer (1995) Agnosticism About Neuron-Level Functionalism; Bringsjord, Selmer (1995) Agnosticism Re-Visited; Bringsjord, Selmer (1995) Are Computers Automata?; Bringsjord, Selmer (1995) Computationalism is Doomed, and We Can Come to Know it; Bringsjord, Selmer (1995) Why Didn't Evolution Produce Turing Test-Passing Zombies?; Bringsjord, Selmer (1996) Artificial Intelligence and the Cyberiad Test; Brown, Marina and O'Rourke, Joseph (1994) Agnosticism About the Arbitrary Realization Argument; Costa, Luciano da Fontoura (1995) Is There More to Personhood Than Computing Theory and Logico-Mathematics?; Hobbs, Jesse (1995) Creating Computer Persons: More Likely Irrational Than Impossible; Korb, Kevin B. (1995) Persons and Things: Massing, Walter (1995) Metaphysical Windmills in Robotland; Mulhauser, Gregory R. (1995) What Philosophical Rigour Can and Can't be; O'Rourke, Marina Brown and Joseph (1995) Agnosticism Revisited; Rickert, Neil W. (1995) A Computer is Not an Automaton; Scholl, Brian (1994) Intuitions, Agnosticism, and Conscious Robots и др.

«вообще», а есть, к примеру, низкоуровневый функционализм, биологический функционализм и пр.

3. Проект ИЛ *реализуем*, так как: 1) человеческая личность – это всё же автомат, но неклассического типа, т.е. представимый не машиной Тьюринга, а иными, квазиалгоритмическими правилами обработки вычислительных элементов; 2) компьютер вовсе не автомат, и кто считает противное – тому не место в проблематике компьютерного сознания; 3) компьютерная система должна тестироваться природой, а не человеком.

4. Проект ИЛ *значим* вне зависимости от возможностей его реализации: 1) будет ли создана ИЛ либо не будет создана – не принципиально, главное, что проекты ИЛ выявляют механистические аспекты человеческого сознания; 2) проектирование ИЛ проясняет роль и функции сознания в жизни человека путём постановки вопроса о зомби – о созданиях, не обладающих сознанием, но имеющих все поведенческие способности человека.

Искусственная личность – экспертная система

Примерно в те же годы, в середине 1990-х годов, в нашей стране отмечался всплеск интереса к проблеме искусственной личности. По заказу Министерства обороны РФ (27 ЦНИИ МО, руководитель – проф. В.В. Деев) были организованы междисциплинарные исследования по внедрению нетрадиционных информационных технологий в системы принятия решений. Автор как программист был в эпицентре этой интереснейшей исследовательской работы. Проект базировался на мероприятиях по моделированию «смысла» и метафорически рассматривался как этап эволюции интеллектуальных систем. Если традиционная компьютерная технология характеризуется формулой «*данные+алгоритм*», технология искусственного интеллекта – «*знания+эвристика (квазиалгоритм)*», то проект ИЛ задаётся формулой – «*смысл+понимание*». Учитывая крупные наработки отечественных специалистов в области построения экспертных систем, в частности в семиотическом моделировании способов представления «знаний», в основу рассматриваемого проекта ИЛ была положена экспертная методология. В её рамках специальные методы анализа процесса «понимания» обеспечивают условия экспликации «смысла», оснований принимаемых решений. Модель

«смысла» фиксирует эти «траектории» в определенным образом кодифицированных массивах информации.

Проблеме моделирования «смысла» посвящено много работ. Среди них в отечественной литературе выделяются следующие модели: онтологическая, культурологическая, психологическая, лингвистическая, герменевтическая, риторическая, дискурсивная, семиотическая, поэтическая, иконографическая, эстетическая, зотерическая, математическая, инженерная, синергетическая, неклассическая и пр. [Алексеев, 2005–4]. Все они так или иначе сформировались исходя из *контекстуального подхода*, согласно которому «смысл» – это обладающий параметрами системного единства контекст формализованных «знаний» экспертов. Обще теоретической основой этого подхода являются классические модели «смысла» Г.Фреге (смысл – это выражаемый знаком способ означивания значения) и её модификаций Б. Расселом и Л. Витгенштейном. В анализируемом проекте ИЛ также использовались идеи теории репрезентации М. Вартовского. В ней в систематическом единстве рассматривается модель, способы её построения и способы её интерпретации. Эта связь обеспечивает динамику наполнения «смыслового объема» возможными «смысловыми траекториями» в ходе согласования метода и предмета репрезентации. Эксперт как бы погружается в квазиалгоритм построения модели модели.

Технически модель «смысла» была выполнена как оболочка над системой представления «знаний». «Смыслообразующими» компонентами выступали категориальные коды «личность», «значение», «смысл», «ценность», «понятие», «действие», «роль», «норма», «умение» и ряд других. Конкретизация и экземплификация этих кодов осуществлялась в концептуальной парадигме теоретико-деятельностного подхода. Эксперт должен соотносить полученные «смысловые траектории» с конкретными формализованными «знаниями», непосредственно задействованными в модели принятия решения. Название проекта – «искусственная личность» – оправдывается тем, что «знания» о принимаемых решениях индексируются персонологическими параметрами.

В принципе, удовлетворяется второй критерий ИЛ – пользователь получает некоторые ответы на запросы этического, эстетического, политического, идеологического и т.п. характера и реализуется персонологическая часть теста Тьюринга. Первый критерий ИЛ не уместен,

хотя разработка проекта сопровождалась спекуляциями на тему «социальной личности» в стиле философии космизма, персонализма, соборности с оглядкой на социологический реализм – о первичности социального над личностным. То есть при построении смыслового объема экспертная система становится «квазисознательной», и эта форма сознания представляет срез «общественного сознания». Конечно, такие суждения не достойны критики. Ещё более «страшной» методологией являлась концепция так называемого «ноокосмизма» – невероятная смесь из псевдофилософских фантазий об интеллектуализации космоса посредством технологий. Разработчик «концепции» – генерал А.Ю. Савин, он же самый главный заказчик проекта. О ноокосмологических «детях Сатурна» мы в нашей работе упоминать не будем.

Апробация экспертного подхода к проектированию ИЛ, осуществлённая в ходе опытной эксплуатации макета, выявила ряд неудач: методологических, организационных, теоретических и реализационных. Однако был достигнут *социокультурный эффект*: предложенный инструментарий вынуждал специалистов осуществлять рефлекссию над «знаниями», эксплицировать «смыслы» собственных решений и представлять их для intersubjectивного обсуждения.

Расширение поля междисциплинарных исследований в проекте искусственной личности

Очевидно, что робототехнический и экспертный проекты ИЛ дополняют друг друга в плане расширения возможностей интеллектуальных информационных технологий. Если первый декларирует:

Искусственная личность – это робот, наделённый маханизмами «квазисознания»

то второй утверждает:

Искусственная личность – это экспертная система, оборудованная механизмами работы со «смыслом»

На наш взгляд, экспертный вариант проекта ИЛ по порядку и по значимости должен предшествовать робототехническому – у него больше шансов пройти вербально-коммуникативный тест Тьюринга. Претензии робототехнического проекта «перепрыгнуть» через задаваемые человеко-соразмерные «смыслы» и «значения» к неким лингвистическим «сущно-

стям» искусственных агентов очевидно поспешны. Необходима предварительная дефиниция персонологических параметров в рамках экспертного подхода, и лишь затем – изучение того, как робот автономно либо в окружении колонии роботов способен, например, продуцировать «коммуникацию», «волеизъявление», «моральные императивы», «религиозные верования» и пр.

Как экспертный, так и робототехнический проекты повышают роль и усиливают значимость междисциплинарного подхода. Собственно и проект искусственного интеллекта уже не мыслим вне поля междисциплинарных исследований психологов, логиков, математиков, лингвистов, нейрофизиологов и др. Проект ИИ значительно расширяет поле исследований, вовлекая специалистов социальных и гуманитарных наук – социологов, политологов, экономистов, искусствоведов, правоведов и др. В нашем проекте ИИ участвовали даже представители паранауки, хотя от них толку было мало. Не специалисты в области естественных и технических наук приоритетны в проекте ИИ. Исключительно важна роль экспертов в области общественных и гуманитарных наук, именно они вводят в предмет конструирования «смыслы», «ценности», «нормы», «идеалы» и прочие составляющие духовной жизни. Требование активной эксплуатации «знаний» социокультурного содержания подчёркивает антропоцентристскую фокусировку целей развития сложных систем, что согласовывается с постнеклассическим ориентиром их построения.

Однако представим утопический сценарий – все лучшие умы страны брошены на реализацию проекта ИИ, выполнены все необходимые методологические работы по коммуникации, координации, интеграции научного сообщества. Более того, со стороны государства получена безграничная финансово-экономическая и правовая поддержка («За искусственную личность платить надо!» – подчёркивал Д. Деннетт [Dennett, 1994-1; 1995-1]). Будет ли реализован проект искусственной личности? Возможно, но для этого надо преодолеть ряд методологических трудностей, связанных со сложностью предмета исследований. Эти трудности на сегодняшний день представляются не только практически, но и теоретически непреодолимыми. Это – проблемы, инспири-

рованные философией сознания. Робототехнический проект явно сталкивается с проблемами сознания, так как антропоморфность робота – первейший кандидат на присвоение ему в условиях тьюрингового тестирования ментальных феноменов – что нужно сделать для того, чтобы компьютер робота функционально уподоблялся мозгу человека? Экспертный подход с проблемой сознания сталкивается неявно – что нужно, чтобы сложная самоорганизующаяся система стала самостоятельно продуцировать «данные», «знания», «смыслы», получающие автономную внутрисистемную интенциональную продукцию, трансляцию и интерпретацию?

Тест Тьюринга, разработанный для прохождения этих проектов ИИ, должен включать в качестве оцениваемых параметров сознательные способности и свойства. К этим способностям, как минимум, относятся: 1) «ощущение» (sentience), т.е. восприятие мира и реагирование на него в соответствии с различными видами перцептивно-эффекторных возможностей; 2) «бодрствование» (wakefulness), т.е. способность фактически реализовывать некоторые возможности, а не только обладать диспозициями к их осуществлению; 3) «самосознание» (self-consciousness), т.е. не только осведомлённость о чём-то, но и осведомлённость о своей осведомлённости; 4) «аспектуальность» (what it is like), т.е. способность воспринимать мир в соотношении с собственной позицией «бытия» или, иначе, восприятие среды в соответствии с ответом на вопрос «каково быть?» (например, каково быть летучей мышью, роботом, экспертной системой?).

Помимо демонстрации вышеперечисленных способностей, системе можно считать квазисознательной, если судья из теста Тьюринга приписывает ей то, что она: 1) пребывает в состоянии «осведомлённости о» (states one is aware of); 2) обладает квалиа – владеет качественными свойствами конкретной разновидности (ощущает красное, чувствует запах кофе и пр.); 3) пребывает в феноменальном состоянии, которое более структурировано, нежели чем квалиа, и обладает определённой пространственной, темпоральной и концептуальной организацией опыта окружающей среды и опыта самого себя; 4) обладает ак-

цессным сознанием (access consciousness) – т.е. доступом к внутриментальным составляющим при управлении речью и действиями (Н. Блок); 5) обладает нарративным сознанием (narrative consciousness) – т.е. выдаёт серию лингвистических выражений, «высказываемых» с перспективы актуальной или виртуальной самости (Д.Деннетт); 6) обладает интенциональностью – направленностью своего внимания на предмет сознания, активностью опредмечивания окружающей действительности (Д.Серль).

Очевидно, что приписывание ментальных свойств лишь опосредованно зависит от обладания системой параметрами физического подобия, как в робототехническом проекте. Конечно, антропоморфность системы значительно повышает степень доверчивости судьи, тестирующего систему. Однако для демонстрации персонологических способностей и свойств вполне достаточно развитого естественно-языкового интерфейса, который предполагается в экспертном подходе к проекту ИЛ. То есть позиция Д. Лената преобладает над позицией Р. Брукса.

Следует подчеркнуть, что вышеприведённые способности и характеристики «квазисознательности» необходимы, но не достаточны. Несомненно, можно привести длинный список иных характеристик квазисознательного поведения. Проблема полноты сознательных способностей и свойств иначе называется «конъюнктивной проблемой», и ей придаётся большое значение в современных дискуссиях наряду с так наазываемой «дизъюнктивной проблемой» – проблемой множественной реализации сознательных феноменов на различных субстратах.

Методологические проблемы философии сознания принято делить по степени трудности, хотя некоторые это отрицают (см. ниже). Очевидно, что методологическая сложность влияет на поэтапность реализации проекта ИЛ – от решения простых проблем к решению более сложных. Вслед за Д. Чалмерсом выделим трудные и лёгкие проблемы.

Проблемы проекта искусственной личности¹

Легкие проблемы – это те, которые затрагивают только операциональные аспекты сознания. Сознание ограничивается следующими феноменами: реакцией на внешние стимулы, отчетами о ментальных состояниях, интегрированием информации когнитивной системой, фокусом внимания, контролем поведения, обучением.

Трудные проблемы связывают сознание с наличием «субъективного опыта» и самосознания. Ключевой вопрос состоит в объяснении механизмов этих субъективных процессов, как каузально связать с компьютерной архитектурой поведение ИЛ, внешне воспринимаемое судьёй в форме субъективного опыта?

По поводу трудной/легкой проблем имеется ряд мнений.

1. Трудная проблема неразрешима.

У. Джеймс в 1890 г., как бы предваряя данный ряд рассуждений, писал: «Природа в ее непостижимых проектах смешала нас из глины и огня, мозга и знания, эти две вещи существуют, несомненно, вместе и определяют существо друг друга, но как или почему, никакой смертный никогда не узнает» (цит. по [Blackmore, 2003]).

Т. Нагель (1994 г.) высказался следующим образом: «Проблема субъективности сложна и безнадежна. Мало того, что мы не имеем никакого решения, – у нас нет даже намёка на то, как ментальное объяснить физическим» [Nagel, 1974; Wider, 1989].

К. Макджинн (1999 г.) считал, что имеющемуся способу изучения ментального и мозга непреодолимо присущ дуализм: «Вы можете изучать ваши состояния сознания, однако никогда не обнаружите нейроны, синапсы и прочее. Вы можете изучать мозг любого человека от рассвета до заката, однако так и не выявите сознание, которое очевидно для того, мозг которого так грубо рассматривается... Мы «познавательльно закрыты» для трудной проблемы – как собака познавательльно закрыта для чтения газеты или понимания поэзии. Как бы собака не пыталась, она не будет способна к математике хозяина, её мозг просто-напросто не так устроен. Подобным образом человеческий интеллект не предназначен для того, чтобы понять сознание. Мы можем научиться нервным

¹ В классификации проблем и подходов к их решению мы частично используем работу С. Блэкмор в [Blackmore, 2003]

корреляциям сознательных состояний (это – легкая проблема), однако не сможем понять, как физическое влечёт сознание» [McGinn, 1999].

С. Пинкер (1997 г.) заявил: «Мы можем бесконечно долго пытаться понять то, как работает ментальное, но понимание этого навсегда спрячено за пределами нашего концептуального горизонта» [Blackmore, P. 32]).

Позицию Т. Нагеля, К. Макджинна и С. Пинкера следует назвать *неомистическим подходом* – они считают, что проблема сознания на самом деле не разрешима, сознание – непостижимая тайна.

2. *Трудная проблема разрешима при выработке нового понимания вселенной.*

Д. Чалмерс требует создания новой теории информации и переосмысления с её позиций «двухаспектной теории информации», согласно которой информация имеет два аспекта – физический и феноменальный. Поэтому сознательный опыт – это, с одной стороны, некоторый аспект информационного состояния, а с другой стороны – аспект, заключённый в физической организации мозга.

К. Клэйрк (1995 г.) требует пересмотра фундаментальной физики. Ментальное подобно некоторым явлениям в квантовой физике. Сознание и квантовый эффект – это субъективные и объективные аспекты представления одного и того же.

Р. Пенроуз (1989, 1996 гг.): сознание зависит от неалгоритмических процессов – т.е. процессов, которые не могут быть выполнены цифровым компьютером или вычислены с использованием алгоритмических процедур. Требуется кардинальный пересмотр отношения ментальное-нейрофизиологическое с квантово-теоретических позиций. Сознательный опыт следует трактовать как качество, обусловленное квантовой последовательностью в микроканальцах нейронов.

Позицию данных авторов можно назвать *неопанпсихическим подходом*, онтологизирующим сознание с позиции атрибутивной трактовки информации. Особенно это характерно для Д. Чалмерса. Однако «подобные взгляды чрезмерно автономизируют категории онтологического и гносеологического, что приводит к ряду теоретических неопределённостей и противоречий», – полагает Д.И. Дубровский [Дубровский, 2002]. То есть проблема усложняется, а не упрощается.

3. *Следует заниматься только легкими проблемами.*

Ф. Крик, А. Сломан, Дж. Маккарти и большинство когнитивных учёных считают, что изучение сознания – не философская, а научная проблема. Мы приблизимся к более полному пониманию проблемы сознания, если начнём с чего-то простого, например с теории визуального закрепления, описывающей синхронизацию электромагнитных колебаний для объяснения того, как различные признаки воспринятого объекта связываются в единство в процессе создания перцептивного целого [Crick, 1994].

4. Нет никакой трудной проблемы.

К. О'Хара и Т. Скатт (1996 г.) в специальной работе «Нет никакой трудной проблемы сознания» [O'Hara, Scut, 1996] приводят три причины, чтобы игнорировать трудную проблему: 1) нам известно, что такое легкие проблемы, с них и следует начинать; 2) решения легких проблем изменяют наше понимание трудной проблемы; таким образом, попытки решить трудную проблему сейчас преждевременны; 3) решение трудной проблемы было бы полезно, если мы могли бы распознать ее как таковую: на данный момент проблема не достаточно хорошо понимается и неясно, какого рода трудность предстаёт перед исследователем.

П. Черчленд (1996 г.) считает, что трудная проблема неверно понята, так как мы не можем заранее предсказать, какие проблемы окажутся легкими, а какие трудными. Откуда мы знаем, что объяснение субъективности более трудно, чем объяснение «легких» проблем? Достаточно ли хорошо определены «трудные» вещи (квалиа, субъективный опыт), чтобы их можно было как-то идентифицировать в роли трудных. Например, мысли – это квалиа или нет? [Churchland, 1996]

Д. Деннетт (1994 г.) считает так: «Даже если бы все легкие проблемы воспроизводства, развития, роста и метаболизма были бы решены, всё равно осталась бы действительно трудная проблема: что такое собственно жизнь. Также и деление проблемы сознания на «легкие» и «трудные» – это отвлекающий фактор, генератор иллюзии» (1996 г.). Например, согласно Деннетту, нет такой вещи, как «феноменальное». Не потому, что отрицается сознание у людей, а в силу того, что философы неправильно трактуют понятие сознания [Blackmore, 2003, P. 32]).

Позицию этих авторов можно назвать *терминологической* – правильно определите то, что называется «сознанием», и тогда проблемы либо чётко обозначатся, либо вообще исчезнут.

5. *Трудную проблему следует решать.*

Это – *конструктивная* позиция. Так полагает Н. Блок. Однако трудную проблему следует дифференцировать для более чёткого понимания предмета исследования. Поэтому предлагается различать сложную и самую сложную (сложнейшую) проблему¹.

Самая сложная проблема заключается в следующем: 1) у нас нет оснований считать искусственную личность номологически и метафизически невозможной; 2) внешняя функциональная эквивалентность может служить основанием для того, чтобы приписать системе сознание или, напротив, отказать ей в сознании; 3) физическая реализация, существенно отличная от нашей, сама по себе не является основой для убеждённости в отсутствии сознания; 4) у нас нет никакого понятия о том, как доказать наличие или отсутствие сознания у другого существа.

В принципе, самая сложная проблема принимает формы других фундаментальных для современной аналитической философии проблем: 1) «проблемы зомби», которая состоит в обосновании мыслимости и возможности (логической, метафизической, натуральной) бессознательных существ, чья физическая структура, внешнее поведение, функциональная организация ничем не отличаются от структуры, поведения и функций сознательных существ (людей); 2) «проблемы *другого*» – на каком основании другим людям, а также иным биологическим и более широко – физическим – системам можно приписать факт обладания ими сознанием (субъективной реальностью); 3) можно показать, что эти проблемы упираются в *проблему создания теста Тьюринга* на предмет обладания системой субъективной реальностью. Ряд исследователей считает, что последняя проблема (тест Тьюринга на субъективную реальность) при сегодняшнем уровне развития науки, философии и методологии теоретически неразрешима (Д.И. Дубровский) [Дубровский, 2004-1]. Но это невозможно лишь пока, на сегодняшний день, в контексте наличных парадигм сознания – бихевиоризма, физикализма, функционализма.

¹ Для раскрытия позиции Н. Блока используется работа [Вильянуэла, 2006]

Сложная проблема состоит в том, что мы не представляем себе, как феноменальное сознание может быть сведено (редуцировано) к нейрофизиологическим фактам, наблюдаемым при изучении мозговых процессов в связи с субъективно переживаемыми явлениями сознания. Почему нейронная основа некоторого феноменального качества есть нейронная основа именно этого феноменального качества, а не другого феноменального качества или вообще никакого. Проблема объяснения данной связи состоит в «объяснительном пробеле» между нейронной основой некоторого феноменального качества и самим феноменальным качеством. Д. И. Дубровский по-иному интерпретирует подобного рода проблему: «Как объяснить связь явлений сознания (субъективной реальности) с физическими (в частности, мозговыми) процессами, если первым нельзя приписывать физические свойства, а вторые ими по необходимости обладают?»¹

При классификации проблем на сложную и сложнейшую логично предположить, что упущена ещё одна, которую назовём «менее сложной проблемой».

Менее сложная проблема заключается в расшифровке нейродинамических кодов ментальных феноменов сознательных существ. Как полагает Д.И. Дубровский, имеется два вида задач расшифровки кода: 1) «прямая» задача, когда дан кодовый объект и требуется выяснить информацию, которая в нем содержится; 2) «обратная» задача, когда нам дана определенная информация и требуется установить ее носитель и его кодовую организацию. Такая задача является более трудной. Однако обе эти задачи – вполне решаемы. Например, известен метод А.М. Иваницкого, который позволяет по электрической активности мозга определять, какие умственные операции совершает в данный момент человек: мыслит ли он пространственными образами или использует логическое мышление [Иваницкий, 2006]. Здесь прослеживается решение первой задачи. Для решения обратной задачи следует воспользоваться принципом инвариантности информации по отношению к физическим свойствам её носителя (принципом Дубровского). Имеется

¹ См. вопросы, сформулированные Д.И. Дубровским для симпозиума «Сознание и мозг», 30 ноября 2007 г., ИФ РАН, Москва и представленные на сайте НСМИИ РАН: <http://www.scm.aintell.ru/default.asp?p0=107>

возможность, при наличии системы изученных кодовых зависимостей, решить и задачу воспроизводства информации на субстрате, отличном от человеческого мозга, – в коннекционистско-репрезентативных кодовых образованиях искусственной личности.

Демонстрация таковой возможности выходит за пределы данной работы, так как требует привлечения специального анализа принципа инвариантности информации и, например, сопоставления его с аргументом множественной реализуемости. Это, в свою очередь, влечёт за собой изучение редукционистских/антиредукционистских парадигм реализации ментальных феноменов на различных физических системах – на мозге человека, на мозге других приматов и животных, на «тине» в «черепе» гипотетических марсиан, на кодовых структурах искусственной личности. Более того, такую работу надо осуществлять в рамках функционализма теста Тьюринга. Сейчас достаточно заключить, что междисциплинарное освоение проекта искусственной личности следует начинать с менее сложной задачи:

**Искусственную личность следует разрабатывать совместно с
расшифровкой кода Дубровского**

Однако решение проблемы сознания (сознание-мозг-компьютер) – будь она простая, сложная, менее сложная или какая-то ещё иная *не* представляется главной стратегической целью развития ИИ. Доминирующей проблемой, по мнению автора, является *проблема творчества*. И если так, то вся перечисленная выше этапность реализации проекта искусственной личности представляется пустым рассуждением, как и все бесконечные метафизические споры о компьютерной реализации феноменов сознания: неосхоластика философии сознания насчитывает десятки тысяч крупных и мелких работ, как показано в классификации Д. Чалмерса. Для практической философии – философии искусственного интеллекта – они бесполезны и порою вредны. Поэтому далее, в подразд. 6.1 мы наметим стратегию переориентации методологии исследования ИИ с психофизической проблемы на проблему реализации творческих способностей человека.

5.3. Критика мировоззрения: комплексный тест Тьюринга против комплексного тестирования в вузовском образовании

Комплексный тест Тьюринга перспективно использовать для критики современного компьютерно-технологического детерминизма, который означает то, что компьютеринг применим для решения всех проблем человека, не только материально-производственных, но и проблем духовной сферы жизни человека и общества. Автора как преподавателя и как специалиста в области проектирования компьютерных технологий особенно возмущает факт крайне неразумного применения идеи компьютеринга для оценки знаний школьника и студента, авторитета преподавателя и ученого. Будучи гражданином, автор не способен понять, как можно доверять компьютеру-арифмометру выборы политического лидера.

Критическую функцию комплексного теста Тьюринга раскроем на примере современных образовательных технологий. Вернёмся к истокам – к оригинальному тесту. Снова подчеркнем, что статья [Turing 1950] послужила становлению аналитической философии искусственного интеллекта, в рамках которой наиболее цитируема и то, что тест Тьюринга (ТТ), неявно обозначенный в статье, послужил прообразом для ряда разнообразных тестов, раскрывающих особенности мышления, понимания, сознания и других параметров субъективной реальности. А.Тьюринг убеждённо доказывает, что через 50 лет (в 2000 г.) машины будут имитировать интеллект, востребованный для ведения осмысленного разговора между людьми, т.е. «диалоговый интеллект» (Н. Блок). Обсуждая теоретические возможности построения мыслящих машин и предлагая проект для их реализации средствами универсального цифрового компьютера, функционирующего в соответствии с заложенными в него эвристиками, А. Тьюринг полагает, что в «будущем компьютеры будут успешно соперничать с людьми во всех интеллектуальных областях» [Управление, информация, интеллект, 1976, С. 282].

Несмотря на амбициозность своей идеи, А. Тьюринг сегодня бы «в гробу перевернулся», узнав о том, как в нашей стране применяются компьютеры в сфере высшего образования. Они уже не «соперничают» с людьми. Они оценивают людей, предопределяя их профессиональную пригодность, материальный достаток, судьбу. Оценка касается не только интеллекта, но и образованности человека в целом, т.е. сложной

многомерной и многоуровневой системы морально-нравственных, творческо-продуктивных, мотивационно-волевых, эмоционально-чувственных, разумно-рациональных и других свойств и способностей личности. Компьютеры конституируют механистические стереотипы мышления культурного человека российского сообщества, формируют менталитет нашего народа, оставляют неизгладимый след на генофонде.

Очевидны вполне справедливые возражения типа того, что сам по себе компьютер не может оценивать человека. В самом деле, если проследить всю цепь тестирования, то мы обнаружим, что в конечном итоге оценивает не компьютер, а человек (судья), используя для этого вспомогательное средство – компьютер. Это так. Но если последовать ещё далее по этой цепочке, то мы увидим, что судья в условиях компьютерного тестирования сам становится своего рода тьюринговым «человеком-компьютером».

Дефиниция «человека-компьютера»

В оригинальной версии А. Тьюринга применяется другой термин – «human computer», «человеческий компьютер», т.е. та сфера в сознании человека, которая отвечает за алгоритмическое поведение. Если осуществить индуктивное обобщение абстрактной способности человеческой личности к осуществлению вычислительных операций, то данная сфера становится определяющим персонологическим параметром. Отчужденный, одномерный человек информационного общества при присвоенной или приданной власти оценивать другого человека, т.е. став судьёй, переносит стереотипы своей калькулирующей рациональности на оцениваемого. Нет особой разницы в том, воплощены ли рассуждения судьи программно-информационными средствами информационных технологий либо реализованы нейро-мозговыми коррелятами его сознания. Функции человека-компьютера и функции компьютера неотличимы. Для обоснования данного утверждения целесообразно сослаться на упоминавшуюся ранее статью С. Харнада «Разумы, машины и Тьюринг: неотличимость неотличимостей» [Harnad, 2001]. Эта работа заслуживает особого внимания в связи с исследованием ментальности людей эпохи электронной культуры. Мы же примем в качестве исходного положение о том, что именно компьютер (человек-компьютер, судья-компьютер), а не человек оценивает человека в си-

стеме компьютерного тестирования. Причём тестируется человек-компьютер. Это верно при принятии тезиса об отчуждённости вычислительной способности человека.

Применим КТТ для анализа концептуальной возможности применения компьютерных технологий тестирования в современной системе высшего образования, которые иногда обозначаются как «e-learning», «инновационные образовательные технологии», «информационные образовательные технологии» и пр. На начало 2009 г. в сфере высшего образования существовали следующие образовательные технологии в масштабах всей страны.

Комплексное компьютерное тестирование в вузовском образовании

Единый государственный экзамен (ЕГЭ). Начиная с 2009 г. является основной (!) формой итоговой государственной аттестации в школе для всех (!) выпускников школ Российской Федерации¹.

Федеральный Интернет-экзамен в сфере профессионального образования (ФЭПО). Проводится в форме компьютерного тестирования студентов и направлен на проверку выполнения требований Государственных образовательных стандартов профессионального образования. Целью ФЭПО является формирование единых требований к оценке качества подготовки специалистов².

Дистанционные образовательные технологии (ДОТ). Непрерывно сопровождают ряд преподаваемых в вузе дисциплин. Это – территориально и тематически распределённые способы и средства преобразования информации, которая составляет содержание учебного процесса, сформированного на основе общего информационно-коммуникационного поля между преподавателями и учащимися. Преподаватели, как правило, выступают в роли виртуальных персоналий, представленных своими научными трудами, учебными и методическими пособиями, тестами. Если на заре развития ДОТ заявлялось лишь об их преимуществах в области вербальной коммуникации (оператив-

¹ См. официальный информационный портал Единого Государственного экзамена: <http://www.ege.edu.ru/>

² См. официальный сайт ФЭПО: <http://www.fepo.ru/>

ность, адресность, преодоление территориальных границ, обширность аудитории и пр.) и подвергалась сомнению возможность преподавания дисциплин, связанных с формированием сложных навыков и умений на основе демонстрации (например, хирургия, опасная техника), то сегодня в условиях развития мультимедийных технологий, искусственного интеллекта, виртуальной реальности и пр. преодолены границы и невербальной коммуникации – сейчас мастер может передавать опыт (например, трепанации черепа), находясь от учеников на расстоянии многих тысяч километров. Вне всякого сомнения, при условии умелого применения, ДОТ способны качественно изменить образовательные возможности¹.

Как мы видим, на этих трех компьютерных технологиях – ЕГЭ, ФЭПО и ДОТ – основана оценка образовательного процесса в вузе. Современный вуз полностью и на всех этапах обучения охвачен компьютерным тестированием. Если первая технология предваряет образовательный процесс в вузе, вторая – время от времени вторгается в него для промежуточной оценки знаний, то третья технология инкорпорирована в учебный процесс. По сути, такие технологии следует считать разновидностями компьютерных вирусов, в широком смысле трактовки последних, – как деятельность человека посредством компьютерно-технологических средств по внедрению в социокультурное пространство для регулярного воспроизводства и трансляции различного рода идей. Подобного рода идея культурных вирусов не нова. Она была выдвинута в теории генно-культурной эволюции Р. Докинса и обрела вычислительные (компьютерно-рационалистские – от сл. «computationalism»)

¹ Сегодня, однако, ДОТ разрабатываются крайне некомпетентно. Пример из личного опыта. Автор решил пройти тест в рамках ДОТ по дисциплине «Философия» в одном из коммерческих «международных» вузов, где преподавал данную дисциплину студентам очного отделения. К его немалому удивлению, имея золотую медаль философского факультета МГУ им. М.В. Ломоносова, кандидатскую степень по философии, многолетний опыт преподавания данной дисциплины, он с трудом вытянул на «удовлетворительно». Правда, вопросы ДОТ лежали вне сферы философского знания и представляли чудесный конгломерат избитых фраз, неизвестных персоналий и цитат из истории, культурологии, политологии, социологии и других дисциплин, но только не из философии.

очертания в теории сознания, предлагаемой Д. Деннеттом и его последователями¹.

Основными участниками технологического процесса тестирования являются два класса отчуждённых людей: преподаватели-компьютеры и студенты-компьютеры. Тестовые вопросы имеют логическую форму простейших логических операций закрытой нестрогой дизъюнкции и закрытой конъюнкции. Крайне редко применяются более сложные логические операции, например, открытый вопрос. Логическая (и интеллектуальная) насыщенность тестов технологически не выгодна, требует дополнительных затрат. В данной статье нет возможности исследовать политические, финансовые, экономические и прочие преимущества и недостатки ЕГЭ, ФЭПО, ДОТ, например, изучить «жизненно важные» вопросы эффективности компьютерных образовательных технологий в борьбе с коррупцией в вузах. Задача достаточно узкая – осуществить концептуальный анализ собственной проблемы образования – проблему оценки того, насколько учащийся усвоил учебный материал и понял ли его вообще. Аргументация направлена на защиту интуитивно очевидного положения о том, что современные компьютерно-коммуникативные новации не решают ключевую проблему образовательного процесса – проблему *понимания* учащимся изучаемого материала. Тем более они не дают возможности понять со стороны оценивающего (судьи), обладает ли учащийся пониманием или не обладает. Они вредят образовательному процессу, так как внедряются слишком поспешно, широко и глубоко. Они ограничивают свободное творчество учащегося узкими рамками механистических калькуляций, ведь приобретение нового знания сущностно творческий процесс. В силу этого системы компьютерного тестирования отнюдь не относятся к инновационным технологиям. Это – архаичные методы старой педагогики, черты которой ярко охарактеризованы в начале 20 в. одним из основателей отечественной философии образования С.И. Гессеном, указывающим на то, что новая педагогика ориентирована на пробуждение творческого начала учащегося, в то время как старая педагогика стремилась оценивать его с позиции механистической зубрёжки (Гессен С.И. Основы педагогики. Введение в прикладную философию / Отв.

¹ См. анализ работ Д. Деннетта, осуществлённый в [Юлина, 2004]

ред. и сост. П.В. Алексеев – М.: «Школа-Пресс», 1995. – 448 с.). Сегодня дряхлая педагогика обрела компьютерную мощь. Поэтому больно отчаяния звучат слова С. Капицы: «Россию превращают в страну дураков... ЕГЭ, который сегодня используют, не может дать объективной картины знаний школьника. Он построен лишь на знании или незнании фактов. Но факты далеко не всё!»).

Следует обратить внимание на фундаментальный труд, направленный на всестороннюю критику ЕГЭ, который был предложен недавно доктором философских наук, профессором В.В. Мироновым, деканом философского факультета МГУ им. М.В. Ломоносова [Миронов, 2011]. Не предполагая никаких априорных предвзятостей по поводу реформы российского образования, в которой ключевым моментом является аттестация абитуриентов по результатам ЕГЭ, В.В. Миронов, тем не менее, убедительно показывает, что ЕГЭ – это крупный шаг назад в развитии образовательного процесса в России. ЕГЭ способствует коррупции, значительно коммерциализирует образовательный процесс (такого грубого «выжимания» денег из молодых людей нет ни в одной стране мира), кардинально ухудшает креативные способности личности..

Интуитивно очевидна справедливость данных слов. Однако вряд ли они будут убедительны для человека электронной культуры, требующего рационального, конструктивного, компьютеризуемого обоснования любых заявлений. Мировоззренческие и методологические основания данной формы культуры пропитаны идеями аналитической философии в сфере математики, логики, лингвистики, техники, когнитивистики и пр. Необходим не метафизический, а аналитический подход, ориентированный на раскрытие смыслов и значений семиотических комплексов, образующих контекст и контент терминов «образование», «компьютерное тестирование», «понимание» и пр. Исследование следует осуществлять, руководствуясь древнегреческим испытанным девизом – «подобное измерять подобным», или, в современном звучании, «компьютерное проверять компьютерным». Компьютерно-ориентированные мысленные эксперименты – тесты Тьюринга – направлены на решение подобного рода концептуальных задач, а представленные в формате *комплексного теста Тьюринга* (КТТ) предназначены для изучения и решения сложных социокультурных задач современного и перспективного информационного общества. К актуальнейшим задачам относится и заявленная проблема приобретения студента-

ми достойного багажа знаний в условиях компьютерных образовательных технологий. В данной работе не раскрывается полная картина применения всего конгломерата тестов для концептуального анализа комплексного тестирования в системе вузовского образования. Совершенно ясно, что достойное образование немислимо без оценки и самооценки творчества, интуиции, логической корректности, математической точности и пр. Анализ достаточно ограничить теми ключевыми моментами системы компьютерного тестирования образовательного процесса, которым имманентен факт субъективной реальности. На мой взгляд, к ним относятся: 1) феномен понимания тестируемым (студентом) изучаемого материала; 2) феномен понимания тестирующим (преподавателем) факта понимания студентом, т.е. приписывание феномена понимания.

В связи с обозначенными моментами воспользуемся соответственно тестом Сёрля и тестом Ватта. Первый тест анализирует человеческую способность *понимать* в условиях компьютерного тестирования. Второй тест анализирует *приписывание понимания*. Обозначим данную разновидность комплексного теста Тьюринга через КТТ₁, где индекс обозначает номер в серии построения интегральных ТТ. Значение индекса равно 1, так как автор данной статьи не осведомлён о том, что другие исследователи применяли подобного рода подходы к анализу феноменов субъективной реальности. Каждый исследователь вносил свою лепту в анализ различных аспектов компьютерной реализации сознания – мыслить, понимать, сознавать, говорить, воображать, творить и пр. Но в решении социокультурно значимых задач наблюдается очевидный пробел. Его мы постараемся восполнить. Рассмотрим совместно данные мысленные эксперименты.

Российский студент в «Китайской комнате» (тест Сёрля)

Воспользуемся классическим мысленным экспериментом Дж. Серля «Китайская комната». В данном тесте поднимается вопрос о прояснении ситуации с пониманием в условиях компьютерных технологий.¹ Наша трактовка очевидна: студент – это человек-компьютер, погружённый в систему программированного тестирования.

¹ Оригинальный тест Серля и сводка дискуссий см.: [Searle, 1980; Серль, 2006; Cole, 2009]

Принято считать, что у эксперимента Дж. Серля были следующие прообразы: 1) «Мельница Лейбница», представленная в знаменитой «Монадологии» [Лейбниц, 1982, С.415-416] – для понимания (ощущения) недостаточно реализующих механизмов, требуется духовная субстанция; 2) «Бумажная машина» А. Тьюринга (1948 г.), в которой показано, что для игры в шахматы требуется не понимание, а лишь вычисление на множестве шахматных комбинаций; 3) «Китайская нация» Н. Блока (1978 г.) – между нейродинамическими состояниями мозга и ментальными способностями (боль, понимание) корреляция не наблюдается.

Дж. Сёрль идёт далее. Суть его эксперимента в следующем. Разыгрывается два сценария. Согласно *первому сценарию* в закрытую комнату помещается англоязычный индивид «Сёрль-в-комнате» (сценарий ведётся от первого лица). Он ни слова не понимает по-китайски. Однако его снабжают китайским алфавитом (исходными иероглифами), английскими инструкциями по составлению из алфавита правильно построенных фраз на китайском и таблицей ответов на китайском. Вопросы задаёт китаец. Сёрль-в-комнате манипулирует символами и выдаёт ответы, бессмысленные для него, но понятные китайцу, находящемуся вне комнаты. При этом Сёрль-в-комнате ничего не понимает. Он работает как компьютер, оперируя входной информацией и выдавая информацию в соответствии с определённой программой. *Второй сценарий* мысленного эксперимента проще. Нет никаких инструкций, алфавита, таблиц и прочего. Вопросы теперь задаёт не китаец, а англичанин. Сёрль-в-комнате всё прекрасно понимает и выдает, конечно, осмысленные ответы на английском. То есть в первом сценарии осуществляется изощрённая информационно-технологическая обработка информации тестирования понимания, но *понимания нет*. Тест на понимание не пройден. Во втором сценарии нет никаких информационных технологий, но *понимание есть*. Тест на понимание пройден.

Очевидна аналогия с тестовой оценкой качества студента. Студент снабжается электронными и учебными пособиями, методическими пособиями, учебным планом, набором тестовых вопросов, на контрольную выборку из которых должен ответить. Причём имеются сроки обучения, которые малы и объёмы материала, которые велики. Поэтому освоение учебного материала обычно *сводится к бездумному запоминанию ответов на тестовые вопросы*. То есть происходит, как и в экс-

перименте Сёрля, формальная манипуляция с синтаксическими конструкциями. В итоге – бессмысленное заучивание тестовых ответов, а понимания-то нет!

Какие могут быть возражения и доводы в пользу того, что понимания нет? Обратимся к работам Сёрля и его критиков.

1. Системологическое возражение. В самом деле, тестируемый студент ничего не понимает, он лишь запоминает тестовые альтернативы. Однако он – лишь часть более общей системы образования. Система же в целом тесты проходит и понимает. Индивид в рамках системы не изолирован. Перед ним – мощные интернет-ресурсы, банки данных с электронными учебными пособиями и пр. Понимание невозможно приписать лишь отдельному индивиду, а следует приписать всей системе образования, частью которой он является.

Ответ Серля мог бы быть таким: пусть студент включит в себя все элементы системы, запомнит всю информацию из Интернета, других баз данных и всевозможные схемы правильных ответов на вопросы. Такой индивид будет олицетворять целостную систему. У данной системы невозможно обнаружить таких функций, которые индивид бы не выполнял, и таких ресурсов, которыми бы он не обладал. Но в любом случае он по-прежнему ничего не понимает по тестируемому материалу, ведь в системе нет ничего сверх того, чего не было бы в индивиде. Если не понимает индивид, то и система также не понимает по той причине, что она стала частью индивида. Если студент не понимает, то соединение его с электронными учебниками не поможет ему понять материал. И далее, уже более жёстко – «...если принимается утверждение, что мне присуще понимание на основании наличия у меня входной и выходной информации определённого типа, а также программы между входом и выходом, то все некогнитивные подсистемы становятся когнитивными. Например, можно составить такое описание работы желудка, характеризующее обработку желудком информации, которое позволяет говорить о том, что желудок «понимает» (Цит. по [Серль, 2006, С.9]). Конечно, следует согласиться с Серлем.

2. Возражение с позиции виртуального сознания. Оно перекликается с первым возражением. Пускай студент, реализуя тестовые задания, не обладает пониманием. Однако обстановка в комнате – книги, базы данных, тексты – это причинный носитель функций понимания,

своеобразный оператор «снабжения» индивида пониманием. Ответ примерно такой же, как ответ на первое возражение.

3. *Робототехнический отзыв.* Допустим, что студент – полноценный робот, «живущий» в среде тестовых вопросов. Программа позволяет ему не только получать формальные тестовые символы на входе и выдавать формальные результаты тестирования на выходе, но, помимо этого, воспринимать окружающее, передвигаться, гулять, забивать гвозди, есть, пить – всё, что может быть отражено в тестовых вопросах. Добавляется много возможностей. Например, посредством системы виртуальной реальности студент обучается трепанации черепа и при тестировании как бы самостоятельно участвует в операции. Будет ли такой студент-робот понимать? Здесь способность понимания обусловлена совокупностью самых различных взаимосвязей с окружающим миром. Однако добавление «перцептивных» и «двигательных» способностей не добавляет ничего нового по поводу понимания. Допустим, внутри робота вместо компьютера находится студент. В отличие от первых двух отзывов, ему предоставляется большее количество тестовых вопросов. Без его ведома часть тестов поступает от телевизионной камеры, присоединённой к студенту-роботу, а другая часть тестов, выдаваемых им, управляет двигателями, двигая руками и ногами и прочими частями тела. Однако все действия сводятся к манипулированию формальными символами. Он получает «информацию» от перцептивного аппарата робота и даёт «инструкции» его двигательному аппарату, не зная об этих аппаратах. Он – гомункул робота, но в отличие от традиционных гомункулов, он не ведаёт, что происходит. Он ни о чем не знает помимо правил манипуляции символами. Следует добавить, что студенту-роботу не присущи никакие интенциональные состояния – он просто перемещается в соответствии с программой и электрическими импульсами в проводах. В свою очередь, руководствуясь программой, студент-робот так же не обладает никакими интенциональными переживаниями. Всё, что он делает – это тупо следует тестовым вопросам и выдаёт запомненные в памяти тестовые ответы согласно бихевиористской схеме.

4. *Отзыв с позиции моделирования мозга.* Предположим, что нейрофизиология и психология достигли такого уровня развития, когда ответ на определённый тест означает определённую конкретную нейродинамическую структуру мозга. При правильном ответе, т.е. при

понимании – одна конфигурация, при неверном ответе – непонимании – другая. Тогда тестирование заключается в сканировании мозга студента. Предлагается представить студента, который перебирает не тестовые альтернативы, а управляет системой водопроводных труб, соединённых клапанами. Когда студент проходит тест, он из памяти извлекает, какие клапаны ему открыть, чтобы выдать правильный ответ, а какие – закрыть. Каждое соединение водных отсеков соответствует синапсу в мозгу студента, а система в целом устроена таким образом, что после должной последовательности возбуждений (что, в свою очередь, соответствует должной последовательности изменений положений вентилей) правильный ответ вытекает из труб, соединённых требуемым образом. Однако понимает ли такая «трубопроводная мозговая» система? Где в этой системе понимание? На входе данная система труб принимает тестовые вопросы и, моделируя формальную структуру мозга студента, на выходе выдаёт ответы. Но студент определённо не понимает смысла тестовых вопросов, так же, как не понимают их водопроводные трубы. «Если же вдруг захочется подумать, что к пониманию способна *конъюнкция* человека и водопроводных труб, то следует заметить, что в принципе человек может представить формализованную структуру водопроводных труб и в своём воображении произвести все «нейронные возбуждения» [Серль, 2006, С.20]. Проблема с моделированием мозга в том, что моделируется лишь формализованная последовательность возбуждений нейронов посредством синапсических связей, а важно моделировать причинные зависимости, определяющие отношение «мозг – понимание». Именно причинные зависимости задают возможность интенциональных переживаний. Формальных зависимостей для этого недостаточно. «Данный факт и демонстрирует пример с водопроводными трубами: мы можем полностью отграничить формальные свойства от важных для возможности понимания нейробиологических причинных зависимостей» [Серль, 2006, С.21].

5. Серль приводит ещё один довод – *комплексный*, который заключается в конъюнкции предыдущих. Возникает вопрос – как из нескольких простых альтернатив и несложных операций возникает сложная сущность ментального характера – понимание? Очевидно, что понимание студентом тестовых вопросов без участия интенциональной связи, внешней относительно этих составляющих, невозможно – понимать он не будет. Но заучить тестовую альтернативу *без понимания* за-

ученного материала, т.е. выступая в роли студента-компьютера, он способен.

Человек-компьютер оценивает человека (тест Ватта)

С. Ватт в работе «Наивная психология и инвертированный тест Тьюринга» [Watt, 1996] утверждает, что тест Тьюринга неявно опирается на «наивную психологию» – т.е. на естественные психологические особенности человека понимать и предсказывать поведение других. В условиях ТТ эти особенности начинают играть главную роль в приписывании ментальных качеств исследуемой системе. В начале рассматривается влияние наивно-психологистской установки на тест Тьюринга – как со стороны тестируемой системы, так и со стороны судьи. Затем предлагается и обосновывается инвертированная версия ТТ. В этой версии процессы приписывания ментальных состояний анализируются явным и более непосредственным образом, чем в оригинальном ТТ. Для наших целей важно принять теоретический факт: человек-компьютер, выступая в роли судьи (судьи-компьютера), принципиально не способен приписать понимание студенту. То, что реальный, эмпирический преподаватель (преподаватель-человек), анализируя результаты тестирования, всё-таки предполагает наличие ментальности у студента-компьютера, по условиям компьютерного тестирования исключается – неформализованные оценки преподавателей не учитываются на стадии статистической обработки баллов, набранных студентом.

Тест Ватта (Инвертированный ТТ) в отечественной литературе не нашел должного освещения. Поэтому его следует несколько подробнее разобрать. С. Ватт начинает с критики оригинального ТТ, априори полагая психологическую обусловленность тестовых вопросов и отказываясь от абстрактных логико-математических интерпретаций теста, в которых, например, длина теста представляется неограниченной или не учитываются особенности различения, присущие судье как реальному человеку, а не как человеку-компьютеру, отчуждённому от всей совокупности невычислимых процедур. Оригинальный ТТ выполнен в бихевиорально-операционалистском стиле, т.е. уже изначально предполагает все возможные оценки того, что оценивается. Судью легко ввести в заблуждение – наблюдатель может приписать интеллект совершенно «неразумным» машинам. Короче говоря, тестируемая система проходит тест Тьюринга, если наблюдатель *верит*, что эта система разумна. Ана-

логичное суждение выносится и относительно других психических свойств и способностей. Применительно к нашему случаю – преподаватель должен быть убеждён, что студент понимает, иначе он понимания студенту не припишет.

С. Ватта данное положение не устраивает, и он применяет методы «наивной психологии», базирующиеся на естественной человеческой установке приписывать ментальные состояния как другим, так и себе, т.е. идентифицировать и воспринимать другое и свое сознание. Наивно-психологическая установка – глубоко укоренённая в человека способность, его фундаментальное природное свойство, существенный, внутренний аспект собственно человеческого поведения. Её следует учитывать в тесте. Категории наивной психологии сравнимы с категориями «природной психологии» (natural psychology), в которой психологические способности объясняются в контексте представлений о социобиологической эволюции животных, и «народной психологии» (folk psychology), изучающей психологические проблемы с позиции здравого смысла. Тем не менее, направленность наивной психологии отличается от выше обозначенных форм психологических исследований. Если налицо реальная естественная способность оценивать «другие сознания», в частности, «другое понимание», то эту способность следует подвергнуть тьюринговому тестированию. Требования к ТТ повышаются: 1) он должен включать вопросы исследования наивно-психологической установки; 2) средства его реализации должны обеспечивать тщательную проверку и контроль фактов приписывания судьёй ментальных свойств тестируемой системе.

В свою очередь, факт приписывания понимания зависит как от поведения системы, так и от судьи – подобно понятию красоты, такие понятия, как «интеллект», «понимание», «сознание» зависят от мнения или точки зрения судьи и неразрывны с социокультурным и психологическим контекстом. Наивно-психологическая установка характеризует готовность вообразить систему интеллектуальной, понимающей, творческой и т.п. даже тогда, когда она таковой не является. Приписывание ментального – это также ментальный феномен. Наивно-психологические тесты – это «тесты на ложную веру» (работы Примака, Вудруфа, Виммера, Пернера, Бэрона – Кохена и др.). Цель тестов состоит в оценке способности ребёнка приписывать ментальные свойства различным объектам. В тестах используются куклы. Перед детьми

разыгрывается спектакль. Необходимо, чтобы в конце представления один из персонажей знал о чём-то ложном и чтобы ребенок знал о ложности таких убеждений. По замыслу создателей этих тестов убежденность персонажа в своей правоте подкрепляет убежденность ребёнка и он не способен демаркировать ложное от истинного. Напротив, тесты «ложной веры» устанавливают границу между ментальными состояниями, которые ребёнок приписывает другим и его собственными ментальными состояниями. И здесь следует ключевая идея инвертированного ТТ: следует совместить наивно-психологические тесты с форматом теста Тьюринга, «перевернув» последний, т.е. на место судьи поставить компьютер. Инвертированный ТТ сравнивает способность различения со стороны системы с той способностью различения, которая присуща опытному судье-человеку. Система проходит тест в следующих случаях: 1) она не способна самостоятельно отличить человека от человека; 2) она не способна самостоятельно отличить человека от компьютера (компьютер при этом проходит нормальный тест Тьюринга); 3) однако она способна самостоятельно отличить человека от судьи-компьютера. Судья-компьютер должен быть судим! Оценивать его работу должен квалифицированный судья.

В итоге получается, что вместо сравнения лингвистического поведения тестируемой системы и человека, как это предполагается в оригинальном ТТ, в тесте Ватта оценивается *соответствие между способностью системы приписывать ментальные свойства другим и такой же способностью у человека*. Для «прохождения» инвертированного ТТ тестируемая система должна показать те же самые закономерности и те же ошибки в приписывании ментальных свойств, которые присущи обычному человеку. Эти закономерности и ошибки далее могут быть проанализированы в психологическом исследовании, базирующемся на изучении антропоморфизмов, ведь именно они в большей мере ответственны за приписывание ментального другим системам. Тест Ватта экономичнее стандартного ТТ, который требует теоретически неограниченной длины для оценки всех возможных ситуаций, относительно которых выносится суждение об интеллектуальности системы. В инвертированном ТТ факт приписывания ментального даёт основание судить, что судья-компьютер – человек, а не компьютер, надо эти факты собрать воедино и систематически сравнить со случаями приписывания ментального, как это делает человек. С. Ватт – сторонник му-

ровской концепции оценки интеллектуальности систем [Moog, 1976, P. 249 – 257], согласно которой ТТ не является дефиницией интеллектуальности, он лишь способствует сбору данных для индуктивного вывода об интеллектуальности. Окончательное решение обусловлено социокультурными факторами. Инвертированный ТТ предлагает новый источник индуктивного доказательства интеллектуальности, базирующийся на правилах приписывания ментальности *другим*, акцентируя на исследовании того, как люди различают вещи, обладающие ментальным и не обладающие таковым. Проясняя роль судьи, которая в исходной версии ТТ совершенно непонятна, тест Ватта вводит в процедуру тестирования самого судью (судью-компьютера), у которого оценивается способность приписывания ментального. Наблюдатель становится наблюдаемым. Оценка осуществляется в обоих направлениях – от человека к компьютеру и от компьютера к человеку, а не только в одном (от компьютера к человеку). Устраняется проблема объяснительной асимметрии – одностороннего процесса, – в котором оценивается лишь компьютер со стороны человека, а компьютер человека не оценивает.

Ранее мы уже упоминали ряд мнений по поводу невозможности организации инвертированного тестирования: 1) тесту присуща дурная бесконечность, так как оценивать судью-человека (который оценивает судью-компьютера) с позиции его человечность/компьютерности должен другой судья-человек и т.д. (Р. Френч [French, 1995]); 2) тесту присуща проблема параллелизма, так как невозможно обнаружить момент, когда судья-человек начнёт понимать различия между собой и судьёй-компьютером (С. Брингсйорд [Bringsjord, 1996]). Данная критика и собственно тест Ватта крайне важны для прояснения концептуальных аспектов компьютерного тестирования, а именно проблем, связанных с изучением приписывания студенту феномена понимания.

Нельзя применять компьютерные технологии для оценки знаний школьника, абитуриента, студента, преподавателя

Рассмотрим ряд очевидных заключений, следующих из совмещения форматов двух частных тестов Тьюринга в концептуальных рамках КТТ₁ с целью изучения способности к пониманию и способности оценивать способность понимания в компьютерных образовательных технологиях.

Комплексный тест Тьюринга

В роли мыслительного инструмента тест представляется крайне эффективным для прояснения значений и смыслов слов «интеллект», «понимание», «сознание» и пр. КТТ – это средство различения человеком тех систем, которые обладают свойствами и способностями, обозначаемые этими словами, от тех вещей, которые этими способностями не обладают. КТТ предназначен идентифицировать сложные социокультурные феномены в отличие от оригинального ТТ, роль которого была ограничена «интеллектом».

Комплексный тест обладает мощным эвристическим потенциалом для осуществления концептуального анализа проблемы понимания. Простое комбинирование частных ТТ позволяет создать многообразие различных КТТ, применимых для нашего анализа. Например, тест $КТТ_2 = [\text{тест Сёрля}] + [\text{тест Ватта}] + [\text{тест Кирка – Чалмерса (тест Зомби)}]$ способен прояснить путь дальнейшего развития сценария использования комплексного компьютерного тестирования в отечественном вузовском образовании (об этом сценарии немного ниже). $КТТ_2$ поднимает проблему демаркации сознательных систем от бессознательных в рамках проблемы другого. Для прояснения данной проблемы требуется построение некоторого $КТТ_x$, тестовые вопросы которого позволят решить, обладает ли *другой* сознанием, подобным тому, которым обладаете вы; понимает ли *другой* понятый вами материал; в большей или меньшей степени понимание *другого* соответствует реальности, нежели чем ваше; имеются ли возможности достижения идентичного понимания? В философии сознания превалирует убеждение, что проблема *другого сознания* на сегодняшний день теоретически не разрешима, хотя имеется целый ряд продуктивных концептуальных возможностей её частичного разрешения в будущем. [Дубровский, 2006-4, С. 153 – 171]. *До тех пор, пока не решена проблема другого сознания, ни о каком*

компьютерном тестировании как главной форме аттестации студента не может быть и речи.

Однако ни КТТ, ни тем более средства компьютерного тестирования (как конкретные средства реализации КТТ) не являются дефинициями образованности тестируемого. Результаты компьютерного тестирования – это лишь подспорье в индуктивном рассуждении по поводу интеллектуальности, образованности студента. Окончательное решение об образовании человека обуславливается не техническими, а социокультурными факторами.

Проблемы студента-компьютера

Втискивание студента в узкие рамки тестовых вопросов, бездумное заучивание правильно подогнанных ответов, превалирование программно-технической составляющей коммуникации и формально-манипуляционных методик оценки качества образования в ущерб личному общению с преподавателем приводит к понижению значимости содержания изучаемого материала. Ущемляются смысложизненные, интенциональные и прочие культурно-антропологические компоненты коммуникации, обеспечивающие прочные связи учащегося не только с реальностью предметной области будущей профессиональной деятельности, но и природой, обществом, человеком. На уровне тестирования особенно остро проявляется экзистенциально-коммуникативная ущербность компьютерных образовательных технологий. Экзистенциальная коммуникация – духовная основа общественной жизни. В большей мере она основана на иррациональных составляющих общения, не поддающихся числовой выразимости. Формальная оценка, полученная как результат тестирования и смысложизненная оценка – несовместимые понятия. Формально «понимание» имеется. Фактически – отсутствует.

Проблемы преподавателя-компьютера

Вопросы, поднятые КТТ, адресуют к оценке сложных ментальных феноменов – способности понимать и способности приписывать понимание. Предполагается, что судья должен обладать более мощными способностями, нежели чем тестируемая система, интеллектуальную систему (таковой является студент-человек) невозможно протестировать неинтеллектуальными средствами (человеком-компьютером), хотя обратное возможно. Кто сегодня оценивает результаты тестирования, «судьи кто?». Конечно, преподаватель-человек, однако его статус понижен. Он вынужден обсуждать с коллегами отнюдь не педагогические проблемы, а то, как «натаскивать» школьников и студентов на все эти бесконечные ЕГЭ, ДОТ, ФЭПО¹. Преподаватель-человек стал преподавателем-компьютером! И его работу оценивает другой человек-компьютер – чиновник от различных ведомств, агентств, министерств – одномерный человек, наделённый мощной бюрократической властью. Инвертированный тест Тьюринга запустить не удаётся, везде тупик «вычислимости», отсутствие свободного созидательного творчества.

Проблемы применения компьютерных технологий тестирования вузовского образования

Вне всякого сомнения, допустима генерализация показателей частной компьютерной методики оценки механистических манипуляций с простенькой программкой. Представляют особую важность легализация на государственном уровне и легитимизация в общественном сознании посредством СМИ результатов подобного тестирования. Крайне значимо возведение данных тестирования в ранг основного показателя аттестации знаний, навыков и умений школьника, абитуриента, студента (в недалёком будущем – преподавателя), а также школы, вуза, региона, области и страны. Однако всё это принимается лишь только в том случае, если мы желаем жить в обществе зомби – квазилюдей, обладающих внешне сознательным поведением, но не использующих для этого ни толики ментальности. К сожалению, зомби уже

¹ В 2009 – 2010 гг. учителя ряда московских школ отказывались проверять тетради учеников, так как им за это не платят, их задача – «натаскивание» на ЕГЭ. К сожалению, они правы. По стране наблюдается много случаев подобного рода.

сейчас среди нас, они появились сразу после развала существовавшей ранее системы образования. О зомби не надо много теоретизировать – посмотрите телевизоры, послушайте радио, войдите в интернет, оглянитесь вокруг, рассмотрите себя. Можно, конечно, принять и допустить широкий охват ЕГЭ, ДОТ и прочими компьютерными «инновациями» системы образования в целом. Всё это допустимо. Вот только биологические потомки не простят нам общество российских зомби, если будут способны понимать факт зомбированности. Возможно, что они не будут это осознавать, так как из современной системы вузовского образования директивно исключёно «понимание». Это мы можем достаточно ясно представить на основе анализа нашего КТТ₁, а убедить – на основе перспективного КТТ₂, построение которого автор намерен осуществить в скором времени.

Согласно тесту зомби, студент-компьютер (он же студент-зомби) может выдавать себя за человека, который освоил материал, хотя фактически понимания с его стороны нет. Компьютерные тесты ЕГЭ, ФЭПО, ДОТ позволяют обмануть преподавателя-компьютера. На самом деле студент ничего не понимает и лишь играет в «кости», представляя случайным образом галочки в ответах. Преподаватель-компьютер не понимает, понял ли учебный материал студент либо не понял. Короче говоря, в рамках комплексного компьютерного тестирования никто ничего не понимает, однако имеются научно-обоснованные, сопровождаемые математическими формулами и диаграммами «юридические» основания, чтобы отчитаться перед чиновником-компьютером за процесс понимания. Чиновник-компьютер при этом «зависает» в пространстве калькуляций, которое не пересекается с когнициями преподавателя-человека и студента-человека. Однако он так же представляет цифровой отчёт другому чиновнику-компьютеру по процессу понимания в его ведомстве. И так далее вверх и вширь по бюрократической иерархии. Очевидно, что в контексте ККТ «понимание» и «вычисление» вносят прямо противоположный когнитивный вклад в образование человека.

По всей видимости, ЕГЭ и ФЭПО будут упразднены со сменой политической идеологии, согласно которой система образования служат формированию не личности, но *потребителя*, как заявил вначале один одиозный министр образования и науки, оправдывая введение ЕГЭ. Второй министр вообще не стал оправдывать развал образования

и науки. Полноценно развитая личность – идеал предшествующей идеологии образования – по причине антропологической константности стремления к самосовершенствованию, станет формироваться отдельно от калькулирующего института образования, и к государственно легитимным способам оценки её компетентности будут относиться столь же поверхностно, как супруги, живущие гражданским браком, относятся к документам ЗАГСа. Многие сегодняшние студенты подобным образом относятся к диплому – не как к свидетельству их знаний, но как к типовому бюрократическому документу.

Тестирование в системе образования следует применять в разумных пределах – в основном там, где требуется оперативная оценка поверхностного знания об изучаемом предмете. Необходимо мягкое внедрение компьютерных образовательных технологий, не наносящее ущерба ни личности учащегося, ни личности педагога. Человек – это не животное, реализующее когнитивные условные рефлексy, которые формируются на основе существующих компьютерных образовательных технологий. Богу – богово, человеку – человеческое, животному – животное, компьютеру – компьютерное. Здесь имеются в виду различные антропологические параметры человеческого существа – религиозные, собственно человеческие, биологические и вычислительные. Отчуждать один параметр в ущерб другим нельзя. Надо строго отдавать себе отчёт – когда, зачем и чем оценивать человека. Человек-компьютер оценить другого человека не способен, субъективную реальность калькуляцией сполна не охватить. Однако если снять тезис отчуждения, оставить человека человеком, разумно использующим свои вычислительные способности, реализуемые технологиями, тогда раскрывается свобода компьютерных новаций в сфере образования. Но сегодня рано говорить про отказ от тезиса отчуждения, учитывая текущее состояние искусственного интеллекта, который лишь в футурологической перспективе позволит интеллектуально насытить компьютерное тестирование и в более-менее полной мере уподобить человека-компьютера реальной человеческой личности.

Компьютер может и должен оценивать человека! Но сегодня следует резко ограничить сферу компьютерного тестирования оценкой чисто механистических параметров деятельности. Для внедрения комплексного компьютерного тестирования в вузовское образование необходимо сосредоточить усилия на междисциплинарных исследованиях

искусственного интеллекта. Лишь после тщательной межвузовской и общественной экспертизы компьютерные интеллектуальные системы тестирования можно применять на практике. Результаты работы этих систем будут во многом способствовать *преподавателю-человеку* в принятии решения о компетентности *студента-человека*.

В заключение отметим, что критические исследования э-культуры, проводимые на базе КТТ, как показал пример с э-образованием, обладают следующими несомненными достоинствами. Во-первых, они применимы для всех феноменов э-культуры, в которых позиционируется некоторая когнитивная проблематика, т.е. изучаются индивидуальные и социальные ментальные феномены наряду с инженерным проектом их компьютерной реализации – их имитации, моделирования или фактического воспроизведения. Во-вторых, когнитивные исследования крайне экономичны. Не надо никаких сложных дорогостоящих социологических, политологических, экономических, идеологических и других анализов. Просто следует подумать по поводу очевидных феноменов ментальной жизни в условиях э-культуры. На наш взгляд, надёжным руководством обоснования таких размышлений служат когнитивно-компьютерные мысленные эксперименты, систематически представленные в формате комплексного теста Тьюринга.

Преимущество нейтрально-зомбистской позиция при решении проблемы сознания в электронной культуре

Проблема сознания в электронной культуре претерпевает сложные смысловые деформации, обусловленные стремлением воспроизвести, смоделировать или, по крайней мере, попытаться симитировать феномен сознания средствами электронных технологий. Семантика термина «сознание» варьируется и зависит от онтологических предпосылок – какую конкретную форму электронной культуры мы рассматриваем.

В э-культуре-1 понятие «сознания» целесообразно рассматривать в традиционном контексте, как это принято в учебниках по философии, культурологии, социологии, психологии и пр. Однако при этом обязательно следует учитывать эргономические коннотаты воздействия на сознание технологий. Например, каким образом влияют на психику ребёнка компьютерные игры. Или каким образом сознание человека, использующего компьютер, приобретает калькулирующий характер (ме-

тафора «человек-компьютер»). Здесь мы имеем бесчисленное множество социологических, психологических, педагогических и прочих исследований влияния электронных технологий на психику человека и на ментальность общества.

В э-культуре-2 понятие «сознания» необходимо дополнять различного рода парапсихологическими репрезентантами, включая концепции эзотерического плана. Например, что значить видеть и ощущать окружающую реальность таким способом, при котором мои глаза находятся, например, на коленной чашечке или в ином другом месте? В виртуалистике подобного рода модели достаточно широко обсуждаются. Возможно, это объясняет тот факт, что основатель отечественной виртуалистики Н.А. Носов, начиная научную деятельность с анализа трансформации сознания у лётчиков, обучающихся на тренажёрах, в конце пути научного творчества перешёл на чистую эзотерику и стал серьезно исследовать такие странные вещи, как «психология ангелов».

В э-культуре-3 понятие «сознания», на мой взгляд, следует рассматривать с нейтральных позиций. В противном случае, принимая сторону зомбистов, мы претендуем лишь на абстрактные эпифеноналистские споры относительно воображаемых клонов-симулякров сознательных существ, своеобразных «номологических бездельников». Как технологически создать бессознательные системы, ведущие себя сознательным образом, и при этом исключить из способа функционирования этих систем хотя бы минимум понятий и представлений о сознании? С инженерной точки зрения это вообще непостижимо.

Вступая в лагерь антизомбистов, мы начинаем строить гигантские проекты расшифровки таинственных нейральных кодов сознания. Это чрезвычайно фундаментальная задача, и предполагаемые результаты её крайне масштабны для всего человечества. Здесь коренится возможность человеческого бессмертия, проникновение в сознание другого. Однако для этого требуется слишком сложная работа, которая под силу не отдельному учёному, научному институту или даже кооперации научных учреждений. Для этой задачи необходимо международное сотрудничество, действующее, по всей видимости, не одно столетие. Поэтому для расшифровки кода Дубровского необходим глобальный научный коллектив.

Нейтрально-зомбистская позиция ориентирована на построение реально действующих проектов. Инженерная проблема сознания в данном случае решается, но при условии риторико-метафорического отношения к понятию «сознания», реализуемого в системах электронных технологий. Чисто искусственные системы и даже симбиотические системы (совмещающие естественные и искусственные «механизмы») никогда не будут обладать «сознанием» в традиционном его понимании – всё это лишь имитации и метафорические эквиваленции «как бы» сознательной деятельности.

Автор придерживается позиции нейтрального зомбиста. Данная позиция позволяет особенно внимательно не задумываться над метафизическими проблемами когнитивно-компьютерной науки и призывает непосредственно переходить к рассмотрению наиболее важной роли, которую выполняет комплексный ТТ в философии искусственного интеллекта – к конструирующей функции. Данная функция служит целью обсуждения проектов, которые, собственно, способны обеспечивать реализацию когнитивных феноменов – осуществлять их имитацию, моделирование и репродуцирование компьютерными средствами.

ГЛАВА 6. КОНСТРУИРУЮЩАЯ ФУНКЦИЯ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

6.1. Конструирование философской методологии: обоснование стратегической роли проблемы «творчество-компьютер»

Дефиниция проблемы творчества в исследованиях искусственного интеллекта

Проблема творчества в традиционной культуре – это постановка вопросов и поиск ответов следующего содержания. Что такое творчество? Как оно возможно? Какую роль играет творчество в жизни человека, общества, в мире в целом? Как отличить творческие действия от рутинных? Как оценить социокультурную значимость творческой деятельности? Какова его природа, состав, структура? Можно ли его описать, проникнуть в сущность, объяснить? Какие методы применимы для понимания феномена творчества? Что такое творческий артефакт? Каковы критерии оценки степени его оригинальности, неповторимости, социальной значимости и востребованности. Где истоки творчества, и каковы способы его интенсификации? Как идентифицировать творческого субъекта? Что такое гениальность? Как найти гения в наличных исторических условиях?

Список традиционных вопросов по проблематике творчества представляется безграничным (см., например, одну из последних работ по философии творчества В.И. Самохваловой [Самохвалова, 2008]). Эти и другие вопросы не имеют однозначных решений, исторически вариативны, социокультурно и субъективно релятивны.

В современной, электронной культуре (э-культуре) проблематика творчества изменяется относительно понятий традиционной культуры. С одной стороны, мы видим расширение понятия, инспирированное технологическими новациями. С другой стороны, ориентир на компьютерные формы реализации творческих явлений сужает понятие, отбрасывая собственно человеческие, экзистенциальные представления о творчестве. Модификация термина «творчество», по сути нарушающая основной логический закон тождества понятия в процессе рассуждения ($a = a$), не мешает изучение данного явления перенести в сферу исследований ИИ и поднять вполне законно интригующий вопрос – можно

ли феномен творчества реализовать посредством электронной технологии, где легальность реализации обеспечивается тьюринговым тестированием? Проблема творчества, как и проблема сознания, становится задачей инженерного дела и целью технологии искусственного интеллекта (ИИ). Таким образом, в первом приближении, *проблема творчества в ИИ – как, чем и зачем осуществлять компьютерную реализацию феномена творческой деятельности – его имитацию, моделирование и/или воспроизведение*. Такая формулировка несколько ограничивает свободу представлений о творчестве и требует дополнения. Дело в том, что имитация, моделирование, воспроизведение – это степени градации онтологических притязаний компьютерной технологии по замене естественного искусственным. При *имитации* понятие интеллектуальности и, соответственно, креативности компьютерной системы, как правило, имеет релятивный статус – интеллектуальные способности, включая творчество, приписываются компьютеру исходя во многом из оценочных убеждений субъекта (наблюдателя, судьи), его пропозициональной установки. Несомненно, технологический инструментарий должен отвечать некоторым приемлемым требованиям организации диалогового взаимодействия человека с компьютером, однако, подчеркнём, «интеллект» во многом специфицирован степенью компетентности судьи, его объективностью, беспристрастностью и другими «юридическими» характеристиками. При компьютерном *моделировании* формулируются условия адекватности модели реальным креативным качествам, свойствам, способностям. При фактическом воспроизведении феноменов творчества – *репродуцировании* – предполагается, что компьютерная система фактически обладает «вдохновениями», «инсайтами», «интуицией», «свободными полётами мысли» и пр. Так или иначе, все эти онтологические притязания выявляются в более широком контексте представлений э-культуры и, как показано выше, зависят от исходной установки исследователей приписывать ментальность х-системам. Данная установка разделяется на следующие позиции: 1) зомбистскую – возможны бессознательные компьютерные системы, обладающие всем спектром ментальных свойств и способностей; 2) антизомбистскую – такие системы невозможны, для реализации сознательного существа на искусственной субстанции требуется репродуцирование ментальности; 3) нейтрально-зомбистскую – имитация и моде-

лирование сознания пригодны для аппроксимативно полной реализации всего выявленного спектра когнитивных функций.

Отметим крайне важную идею, которую привносит проблематика творчества, существенно расширяя сферу ИИ, обновляя традиционные понятия. В компьютерную реализацию когнитивных феноменов включается ещё одна составляющая, которую обозначим термином «креация». *Креация* – это совокупность способностей и свойств компьютерной системы продуцировать феномены сверхантропологического и неантропологического статуса. Сверхантропологические компетенции в некотором смысле очевидны – они превосходят индивидуальные человеческие способности, например, при компьютерной игре в шахматы нешахматиста. С пониманием неантропологических феноменов сложнее – это явления, несоизмеримые с человеческой ментальностью. Человек, в силу родовой предопределенности, не способен оценить способ функционирования компьютера как «творческий» – в человеческом понимании значения слова «творчество», подобно тому, как человек не способен оценить разум животных. Компьютер – это своеобразный Солярис, о разуме которого можно лишь опосредованно догадываться по следам-символам (см. прекрасную феноменологическую ремарку Н.М.Смирновой [Смирнова, 2006]). Однако ряд авторов, изучающих робототехническую проблематику реализации когнитивных феноменов (Дж. Маккарти, А. Сломан, С. Бригсйорд и др.), полагают, что с «ментальностью» компьютерных систем дело обстоит проще, нежели чем с «психикой» других систем, скажем, животных. Мы ведь многое знаем по поводу информационных процессов, протекающих в роботах, в то время как у животных, включая самих себя, об этих процессах можем догадываться только по аналогии, причем часто по аналогии с компьютером. Такая позиция возможна для нейтральных зомбистов.

Таким образом, с учетом вышесказанного, предложим следующую общую формулировку:

Проблема творчества в искусственном интеллекте состоит в том – как, чем и зачем имитировать, моделировать, репродуцировать и креацировать феномены творчества средствами компьютерной технологии

Феноменология компьютерного творчества представляется распределённой среди всех перспективных направлений ИИ. В методоло-

гических исследованиях искусственной жизни изучается преимущественно витальный феномен – как посредством компьютеринга реализовать феномен жизни. В искусственной личности главной выступает проблема компьютерной реализации ментальных феноменов – квалиа, ощущений, восприятий, представлений, понятий, суждений, рассуждений и пр., а также способов «самособираания» этих феноменов до уровня эпистемологической адекватности, эстетического наслаждения, морального вменения, самости, личности, индивидуальности, Я. В искусственном обществе прерогатива исследований принадлежит социально-когнитивной феноменологии. Очевидно, что агенты этого общества связываются в многоагентную систему не путем механистической состыковки «по запросу», выделяющему типовые свойства и методы агентов, но за счет креативных процедур, основанных на менталистских параметрах – репрезентантах «потребностей», «мотивов», «целей», «намерений», «убеждений», «любви», «надежды» и пр. Конечно, здесь востребована проблематика социологического компьютеринга, которая на сегодняшний день находится в зачаточном состоянии – слишком велика разница между общетеоретическими социальными теориями и их когнитивно-компьютеринговыми инженерными приложениями. «Когнитивная культурология» с её представлениями о культуре как некоторой программной управляемой совокупности баз данных и баз знаний для нормального культуролога расценивается как сценарий фильма ужасов. В когнитивной социологии реализации феноменов творчества следует отвести значимую роль, по крайней мере, сравнимую с ролью, которую приписывают «творческой эволюции» в биологии и «творческой деятельности» в социальных науках.

Стратегия развития ИИ: проблема творчества vs проблема сознания

В философии ИИ, осуществляющей мировоззренческую рефлексию и методологическую организацию научной и практической деятельности по построению и применению интеллектуальных систем, крайне важным представляется поиск фундаментных проблем, от решения которых надежно и основательно зависит успех в развитии всех обозначенных направлений ИИ. В отечественной методологии ИИ роль фундаментной проблемы играет «*проблема сознание – мозг*». Проблему обозначил сопредседатель НСМ ИИ РАН проф. Д.И. Дубровский и, как

в силу его авторитета, так и по объективным причинам (в самом деле, проблема сознания стратегически важна для судеб земной цивилизации – в этом автор полностью убеждён), проблема попала на топ-вершину междисциплинарных проблем ИИ. Суть состоит в компьютеринговом воспроизведении информационных зависимостей феноменов сознания от мозговых коррелятов. Однако справедливости ради отметим, что ряд видных исследователей ИИ, более полвека отдавших развитию этой технологии, ни о какой «проблеме сознания» ранее не слышали. Незначительную роль отводил «проблеме сознания» и основатель философии ИИ – А.М. Тьюринг, вскольз парируя контраргументы «от сознания» и «от х-возможностей» от линии защиты идеи компьютерной имитации мышления [Turing, 1950, P. 445 – 450]. Первое возражение полусутоливо снимается исходя из солипсической позиции – по поводу сознания другого человека мы делаем заключение лишь по аналогии со своим собственным сознанием и по большей части приписываем ему феномен сознания в силу вежливых условностей. Точно так же, как мы приписываем сознание другому человеку, мы можем приписать сознание и компьютеру, который оценивается в условиях теста Тьюринга. Второй контраргумент – «от различных возможностей» – А. Тьюринг также посчитал незначительным. Эти возможности, по сути, редуцируются к частным аспектам сознания и к комбинациям этих аспектов. Как и предыдущий, данный аргумент опровергается не совсем серьезно: машине «любить клубнику» надо в той же мере, как и «сиять на конкурсах красоты». Собственно, зачем машине обладать сознанием? Машина должна работать и, в случае с ИИ, решать задачи, которые принято считать собственно интеллектуальными.

В современных инженерно-технологических разработках пока не место компьютерным механизмам, которые реализуют ментальные способности. Сегодня эти модели ничего не значат для эвристического извлечения и кодирования информации, оптимизации и рационализации процессов её репрезентации, хранения, поиска, манипуляции, отображения, передачи, для идентификации значимых проблем и достижения целей сложных динамических систем посредством адаптации к неопределённым условиям среды при нехватке ресурсов, для управления «знаниями».

Приписывание способу функционирования компьютера феноменов «переживания», «восприятия», «понимания» и пр. – это оценка

функционирования *креативных программ* – программ, осуществляющих реализацию творческих актов. Имеется претензия на то, что х-система производит нечто оригинальное – переживание, представление, мысль, значит, есть повод приписать данной х-системе «сознание». Никак иначе. В далёком будущем, когда, возможно, будет расшифрован код *Дубровского* (широко известный в отечественной литературе подход к нейродинамическому кодированию ментальных феноменов), когнитивные модели «сознания» станут фундаментом ИИ, включая основы для формирования более совершенных креативных программ. Но даже в такой перспективе собственная идея развития ИИ будет предполагать необходимость креативного создания новых средств и методов, что означает использование креативных интеллектуальных систем. То есть идея креативности не снимается, даже если в будущем каким-то образом будет решена проблема «сознание – мозг – компьютер».

Для более чёткого позиционирования проблемы сознания и проблемы творчества поиграем в «слова», как это принято в лингво-аналитических исследованиях, проясняющих смысл и значения семиотических комплексов.

1. Для начала допустим, что понятие творчества включает понятие сознание. Для данной позиции ответим на вопрос – собственно, каким образом из массива знаний образуется «со-знание» как целостная интегративная форма совместных «знаний», контент которой воспроизводит квалиативные, феноменальные, интенциональные, позиционные, доступные, нарративные и другие внутренние особенности сознания как субъективной реальности? Эволюционные, эмерджентистские, синергетические, системологические, аутопоэтические и прочие «механизмы», которые приближают нас к объяснению данного феномена, в обыденном словоупотреблении удобно и принято обозначать словом «творческие способности». Таким образом, «творчество» играет роль инфра-понятия, на базе которого возможно объяснение факта объяснения «сознания». Более того, «творчество», особенно в наиболее гениальных формах проявления часто приписывается сфере бессознательного или сверхсознательного. Здесь вновь срабатывает принцип несущестственности сознания Фланагана, который теперь звучит не относительно интеллекта, а относительно феномена творчества: «*Творчеству сознание не нужно*».

2. Изменим денотативные роли терминов. Теперь будем полагать, что творчество – это высшая функция сознательного существа, т.е. концептуально подчиняется «сознанию». При данной раскладке ролей мы видим, что творчество – частный акт сознания, пускай, наиболее значимый. Однако на основании чего можно приписать ментальные параметры х-системе, в частности, компьютерной системе? Лишь только по способу её функционирования, который оценивается как творческий! Оценка сознательной деятельности базируется на приписывании системе творческих способностей, если мы не способны приписать креативность результатам функционирования компьютерной системы, ни о каком приписывании «сознательного» говорить не приходится.

Следовательно, проблематика творчества даже при условии признания того, что данный феномен находится в концептуальном подчинении у феномена «сознание», имеет более важное значение для развития методологии ИИ, нежели чем проблематика сознания – этот феномен можно непосредственно оценить, в отличие от феномена сознания, который познается лишь интроспективно или по аналогии, что является логически некорректным способом аргументирования по поводу наличия/отсутствия сознания и, тем более, его параметров.

Таким образом, исходя из ряда очевидных положений, мы видим некоторую «искусственность» проблемы сознания в контексте исследований искусственного интеллекта. Проблема иллюзорна и не обладает практическим потенциалом даже в силу наших простых рассуждений. Большей значимостью для философских исследований ИИ и по показателям оценки поведения системы, доминирующей проблемой, представляется *проблема творчества*. Собственно от первичного решения данной проблемы зависят успехи ИИ. Остальные проблемы, включая «проблему сознания», представляются вторичными и производными от проблематики компьютерной креативности. Отсюда следует тезис:

В междисциплинарных исследованиях ИИ первичной является проблематика творчества, но не проблема «сознание–мозг–компьютер».

При ответе на основной вопрос методологии исследований ИИ – что первично – проблема сознания или проблема творчества – отметим первичность креативной проблематики. Однако наряду с этим следут указать на аргументативную слабость предложенной выше лингвоана-

литической игры – это всего лишь «игра в слова». Более основательный довод формулируется средствами КТТ.

Проблема творчества в методологии комплексного теста Тьюринга

Комплексный тест Тьюринга позволяет позиционировать проблематику творчества в ИИ и классифицировать направления исследований, исходя из собственных когнитивно-компьютерных понятий, не прибегая к интуициям биологического, психологического, социологического, антропологического и тому подобного дискурса. Для классификационной задачи, которая следует за дефинитной задачей определения когнитивных феноменов, в первую очередь следует отграничить проблематику креативных систем от тех компьютерных систем, в которые не заложены принципы компьютерной реализации творческой деятельности. Как правило, при терминологическом обозначении нетворческих компьютерных систем используются психологические метафоры, характеризующие интеллектуальные отклонения.

Обозначим класс компьютерных систем, которые номинативно обозначаются термином «интеллектуальные» (однако способу функционирования которых сложно или невозможно приписать креативную способность), термином *кретинные системы*. Организация таких систем не включает концепций интеллектуального развития, т.е. самостоятельного прироста знаний. В качестве подкласса кретинных программ можно указать *савантистские системы* – узкоспециализированные системы ИИ, настроенные на конкретную предметную область или под специализированные знания конкретного человека, которого предполагается обыграть в условиях ТТ. Психиатрический синдром «саванта» обозначает психическое отклонение, наблюдаемое, как правило, у аутистов, при котором люди, выдающие гениальные способности в одной узкоспециализированной сфере деятельности (невероятный объем запомненной ими информации, огромная скорость арифметических вычислений и пр.), в других областях крайне ограничены, вплоть до неспособности жить. На наш взгляд, к савантистской программе относится Deep Blue, шахматный суперкомпьютер фирмы IBM, который в 1997 г. одержал победу в матче из 6 партий с чемпионом мира по шахматам Г. Каспаровым.

В контексте КТТ наиболее полно кретинные системы представлены в рассмотренном выше параноидальном тесте (тесте Колби). Па-

параноидальная система – программа, имитирующая способ мышления и общения параноика – человека, страдающего подозрительностью, хорошо обоснованной и стабильной системой сверхценных идей, которая порой выражается бредом. Параноидальные системы – это все существующие системы, построенные в рамках репрезентативной парадигмы. В самом деле, компьютерная онтология этих систем навязывается концепцией программиста и за границы её не выйти, так как нет саморазвивающихся средств, встроенных в программно-информационные средства. Другое дело – системы ИИ, построенные в рамках коннекционистской парадигмы. Здесь Д. Деннетт обращается к идее антропоморфных роботов Р. Брукса, в особенности, к роботу КОГ. Для данного видного исследователя философии ИИ очевидно, что только нейрокомпьютерные системы способны отвечать приемлемым конвенциям интеллекта – «знания» в этих системах «активны» и «свободны».

Мы также полагаем, что кретинные системы не относятся к системам ИИ, хотя часто называются «системы искусственного интеллекта». Например, это упомянутая в тесте Маккарти классическая экспертная система MYCIN, а также системы, комплексно продуцируемые индустрией «динамических экспертных систем реального времени».

Креационистские системы

Потивоположностью кретинных систем, с точки зрения проблематики творчества, являются системы, в которые заложена идея творения. Данный класс программ обозначим термином «креационистские системы» и соответствующее им направление в философии ИИ назовем *креационизмом* (от лат. «творение»). В общем случае данное направление порождено комбинацией идей ТТ с идеями фон Неймана по поводу реализации самовоспроизводящихся машин. Компьютерный креационизм крайне востребован в проектах «искусственное общество» и «искусственный мир». Здесь мы видим своеобразную теологическую окраску исследований ИИ – компьютер как бы творит «бытие» из «ничто», т.е. той субстанции, которая в различных мирах обладает общими инфрасистемными особенностями. Конечно, «богом» в этой схеме выступает человек-программист.

Тест Баресси – креационистский тест Тьюринга

Для изучения проблематики творения наиболее значимый частный ТТ представлен в работе Д. Барреси «Перспективы Кибериады: границы человеческого самопознания в кибернетическую эпоху» [Barresi, 1987], опубликованной в «Журнале теории социального поведения». В истории тьюрингового тестирования данная модификация ТТ получила название «Кибериадный тест Тьюринга». Фоном для рассуждений Дж. Барреси служит серия рассказов польского писателя-фантаста С. Лема, считающего их философскими притчами о технологии будущего, опубликованных с 1964 по 1979 гг. и объединённых общим названием «Кибериада» (отсюда, собственно, и название теста). Дж. Барреси ссылается на работы, которые связаны общей картиной фантастической псевдо-средневековой вселенной, где место людей занимают роботы.

Барреси продолжает философский путь С. Лема, но конкретизирует его рассуждения применительно к ИИ в целом и к ТТ в частности и выдвигает ряд оригинальных философско-прогностических идей как по поводу постчеловеческих ориентиров ИИ, так и по приложению их к проблематике творения в исследованиях ИИ. У Дж. Барреси просматриваются два направления проблематики творения искусственного мира – универсальное и эволюционное. В *универсальном креационизме* изучается компьютерная реализация акта «творения» посредством машины-бога, который «всезнает» о том, каким должен быть «творимый» им мир, т.е. компьютер-творец обладает абсолютным знанием и всеовершенными инструментами для творения «себе подобных по своему образу». В *эволюционном креационизме* за инженерную реализацию отвечает «машина-природа», в которой действуют «законы» мутации и селекции компьютерных средств. Здесь нет претензии на универсальность знаний о творении «третьего мира», искусственная система служит средством самопознания творческих способностей субъекта, творящего «искусственный мир».

Как для универсального, так и для эволюционного креационизма отметим две схемы: горизонтальную и вертикальную. *Горизонтальный креационизм* выдвигает концепцию, согласно которой искусственный мир создается параллельно естественному. Искусственные личности этого мира (роботы, киборги, артилекты и пр.) рождаются и творятся в контекстах, всем нам известным, в нашем мире. Здесь реализуется тезис А. Тьюринга по поводу того, что в будущем компьютеры будут заме-

щать человека во всех сферах общественной жизни [Turing, 1950, Р. 460]. Другое дело – вертикальная схема. *Вертикальный креационизм* совершенно захватывает дух – человек творит искусственный мир искусственных личностей. Далее искусственные личности искусственного мира 1-го порядка творят искусственный мир 2-го порядка и т.д. Несомненно, это крайне важные направления для развития технологического базиса э-культуры, т.е. НБИКС-комплекса.

Отметим, что кибериадный ТТ не имеет репрезентативной интерпретации. *Креационистские системы исключительно нейрокомпьютерные*. Творимый мир в рамках горизонтальной схемы, либо мир, в котором творится искусственная система, представлены в формате огромной нейронной сети. Компьютируемая самость формируется в областях наибольшей коннекционной интенсивности. Однако такой взгляд требует последующего прояснения и выходит за рамки данной работы.

Для прагматически ориентированных «творческих» компьютерных систем ни кретинные, ни креационистские схемы не интересны. Кретинные системы слишком «глупы», креационистские системы слишком фантастичны. Интерес представляют собственно «творческие» системы ИИ, креативные системы.

Креативные системы

В креативной системе компьютерный способ реализации феномена творчества представляется основным принципом функционирования и/или внутренней организации. В философии ИИ проблема создания креативных систем начинается с оригинальных работ А. Тьюринга [Turing, 1950, 1964]. Здесь мы наблюдаем три класса креативных систем: машина-оракул, машина-взрослый, машина-ребёнок.

Предложения и последующие дискуссии по первому классу систем породили направление в философии ИИ, которое мы обозначим как *универсальный креативизм*. Они создаются и функционируют в соответствии с универсально-эпистемологической парадигмой роста «знаний» и детально рассмотрены выше. В этих исследованиях мы видим проект компьютерной системы, творческие способности которой превосходят человеческие в социально-эпистемологическом смысле. Такой компьютер получил название «машина-оракул». Линию обсуждения конструкции «машина-ребёнок» обозначим как *эволюционный креативизм*. Они развиваются в соответствии с эволюционно-

эпистемологической парадигмой роста «знаний». В дискуссиях данного направления верхом совершенства компьютерной системы предполагается программная реализация способа социализации ребёнка – первоочередно овладение речью (в робототехнике – «псевдоречью») и стереотипами поведения окружающих людей (копирование поведения, частично воплощённое в проектах антропоморфных роботов).

Вариант «машина-взрослый» предполагает совмещение универсального и эволюционного подхода в формате общего проекта, и направление исследований креативных систем обозначим *как умеренный креативизм*. Последний вариант – гибок и представляется наиболее перспективным для стратегии построения интеллектуальных компьютерных систем.

Выше данные классы систем были рассмотрены с позиции наличия в инженерной конструкции проекта компьютера эволюционно-эпистемологических механизмов. Проблема творчества, таким образом, представляется феноменологическим аспектом изучения перспектив интеллектуального компьютинга, ЭЭ-механизмы – субстратным аспектом, наличие их в системе концептуально необходимо для того, чтобы данная система могла пройти креативный тест.

Универсальный креативизм

Универсальный креативизм в контексте ФИИ означает построение и использование компьютерной системы, в которую разработчик (эксперт, когнитолог, программист и др.) априори вложил совокупность таких возможностей, которым наблюдатель (судья) готов приписать параметры творческой человеческой деятельности. Универсализм по аналогии с биологическими теориями означает то, что система самостоятельно не развивается, её разработчик всё учел и «все проник». Здесь возможна классификация креативных компьютерных систем по специфике предметных областей (игра в шахматы, сочинение поэм, музыкальных произведений и пр.) Также следует выделить в качестве идеального варианта абсолютно универсальную систему. Компьютинговая мощь её достаточна для имитации любого творческого проявления человеческого гения. Однако возможна ли такая машина? Вспомним то, что ещё Р. Декарт признавал невозможность создания такой абсолютной ма-

шины – он полагал, что машину всегда можно будет легко отличить от человека, если протестировать её в широком тематическом диапазоне.

Сегодня в связи с глобализацией системы управления «знаниями» правоту Декарта оспаривает интернет. На этот счет имеется важный мысленный эксперимент – тест Н. Блока. Данный тест ранее фигурировал для обоснования бесперспективности универсально-эпистемологических механизмов развития «знаний». Здесь мы приведём пример теста Блока с целью раскрытия важности выбора проблематики сознания как стратегического пути развития ИИ.

Тест Блока: подтверждение преимущества проблемы творчества перед проблемой «сознание-мозг-компьютер»

Результаты данного теста двусмысленны, хотя Н. Блок этого не заметил, так как не исследовал креативных параметров тестирования. В этом тесте наблюдатель (судья), приписывающий ментальность х-системе, как и в оригинальном ТТ, не может наблюдать физиологические параметры игроков. Но теперь он играет с людьми всей земли – теми, которые, по крайней мере, выкладывают контент в интернете и реализуют поисковые алгоритмы информационно-аналитических технологий. Предполагается, что на любой запрос наблюдатель может получить ответ – кто-то же должен знать, что ответить. На самом деле это не так, имеются нетривиальные вопросы, на которые некому ответить. Это предполагает либо творческую работу над собой, либо работу креативных специалистов-разработчиков, способных ответить на вопрос, если вообще на него можно ответить.

Н. Блок почему-то не замечает явных способов организации творческой деятельности и действует с метафизических позиций теории сознания, которую он назвал «психофункционализм». Данная разновидность функционализма сегодня считается классической среди англо-американских исследователей философии сознания. Блоковская концепция сознания означает невозможность указать однозначное отображение между входом и выходом системы. Попытка описать такое отношение невозможна без учёта контекста всей совокупности психических свойств и способностей. Это порождает континуум способов преобразования

входных данных в выходные результаты и редукцию исходного процесса к субпроцессам, последних к суб-суб-процессам и т.д. Данная концепция была задумана Н. Блоком для антибихевиористского опровержения оригинального ТТ и, по мнению автора, не позволяет однозначно сконструировать инженерные проектные решения. То есть глобальная универсальная креативная система (типа мыслящего интернета) невозможна. Конечно, это так. Только непонятно, зачем Н. Блоку потребовалась апелляция к метафизике сознания, ведь намного проще неработоспособность универсальной статичной системы объяснить посредством более очевидных положений, исходя из отсутствия в ней *прироста знаний*. Собственно здесь мы находим *подтверждение приоритета проблемы творчества над иллюзорной и схоластической проблемой сознания в исследованиях ИИ*. С другой стороны, находим подтверждение тезиса о важности методологии комплексного теста Тьюринга в философии ИИ.

Чтобы прояснить более интересные варианты креативных систем – эволюционной (машина-ребёнок) и умеренно-универсальной (машина-взрослый), – обратимся к оригинальному ТТ.

Проблематика творчества в оригинальном тесте Тьюринга

Проблема творчества встала перед Тьюрингом при обсуждении основного вопроса философии ИИ – «может ли машина мыслить?». Напомним полемический стандарт, предложенный А. Тьюрингом для ответа на возможные возражения по данному вопросу. К дискуссии относятся следующие темы: теологическое, антисциентистское («голова в песке»), математическое, «аргумент от сознания», «аргумент от х-невозможностей», «возражение леди Лавлейс», «аргумент от непрерывности нейронной системы», «аргумент от неформальности поведения», «аргумент от экстрасенсорного восприятия» [Turing, 1950, P. 443 – 454].

Среди аргументов наибольшую значимость А. Тьюринг приписывает *аргументу Лавлейс* (АЛ) [Turing, 1950, P. 450]. С обсуждения данного аргумента берет отсчет проблема творчества в ИИ и начинается исследование компьютеринга феноменов творчества.

Возражение леди Лавлейс буквально таково: «Аналитическая машина не претендует ни на что оригинальное. Что бы машина не сде-

лала, нам известен способ, как она это сделала» [Turing, 1950, P. 450]. То есть машины не способны к творчеству, так как сами по себе не могут создать ничего нового, они лишь следуют программным инструкциям. Все «творческое» предопределено программистом.

Важность роли, которую А. Тьюринг придавал аргументу Лавлейс, проявляется не только в объеме соответствующей части статьи, но и в том, что предлагается концептуальный проект компьютерной системы, которая способна имитировать и моделировать феномен творчества – нейрокомпьютерный проект машины-ребёнка и гибридный проект машины-взрослого.

Ранее мы выделяли два варианта ответа на возражение А. Лавлейс – репрезентативный и коннекционистский. Коннекционистской парадигме КТТ посвящается оставшаяся часть работы. Поэтому остановимся подробнее на репрезентативном опровержении.

Тест Лавлейс: аргументативная слабость репрезентативной парадигмы креативного компьютера

Репрезентативное опровержение АЛ, предпринятое А. Тьюрингом, на мой взгляд, недостаточно обоснованно, так как все его доводы не удовлетворяют минимуму требований теории аргументации, т.е. логически некорректны.

1. А. Тьюринг перефразировал АЛ и высказал следующее: «чтобы машина ни делала, на самом деле она ничего нового не делает» [Turing, 1950, P. 450]. По поводу этого довода он выдвигает «контраргумент»: «Под солнцем ничего нового не происходит... Всё, что мы считаем оригинальным, на самом деле представляется результатом обучения или следствием известных общих принципов» [Turing, 1950, P. 450].

2. А. Тьюринг вновь перефразировал АЛ: «машины никогда нас не удивят». Теперь он возражает: отнюдь, машины его удивляли очень часто, например, если они сбивались во время работы.

3. Возражение АЛ отводится ещё и потому, что человеческая способность к творчеству непонятна, невозможно найти подлинного автора идеи, так как помимо автора, крайнего на данный момент време-

ни, имеются и другие «авторы», которые его обучили и воспитали. Совершенно не очевидны критерии оригинальности. Даже если не брать во внимание компьютерные сбои, результаты работы машины способны поразить любое воображение.

Несмотря на то что исследователи ТТ часто удивлялись «удивлению Тьюринга», т.е. тому, что компьютерные сбои преподносят сюрпризы, которые расцениваются как акты творчества, следует отметить, что в научно-фантастической литературе и в кинематографе к этим нарушениям относятся вполне «серьёзно». В результате некоторого программного сбоя у роботов, как правило, возникает сознание и самосознание. Здесь играет роль интуиция в вопросе о креативности феномена сознания, о возможности его возникновения за пределами алгоритмической регулярности. Если обратиться к практике программирования, то следует отметить, что ошибка имманентна программе – кто программировал, прекрасно знает, о чем идет речь. На этот счет имеется суровый «закон Мерфи», согласно которому в любой программе, в которой выявлено n ошибок, имеется как минимум $n+1$ ошибка. Так что программы удивляют не из-за сбоя в технических средствах, как это было во времена А. Тьюринга, но и по причине специфики собственно процесса разработки программ. У программистов, которые рассматривают свою профессиональную деятельность как форму творчества, применяется интересный термин «клудж». Это сотканный из разных модулей программный код, который, несмотря ни на что, работает. Сам разработчик не может понять, почему программа работает. Клудж, конечно, удивляет – каким образом вдруг неожиданно начинает работать комплекс, состоящий из набора подпрограмм, каждая из которых не работает в силу несоблюдения соглашений между локальными и глобальными переменными, или, например, отношений между формальными и фактическими параметрами, не говоря уже о других причинах. Чтобы клуджевую программу внедрить, требуется немного – более-менее точно определить условия, при которых комплекс клуджей устойчиво функционирует, и разработать соответствующую документацию.

4. Ещё одну интерпретацию репрезентативного опровержения АЛ следует обозначить как «аргумент от неведения». Проблематика компьютерного творчества изучается А. Тьюрингом на базе компьютера, который подразумевается в контексте категорических суждений А. Лавлейс. Имеется в виду аналитическая машина Ч. Бэббиджа (Babbage's Analytical Engine). Здесь А. Тьюринг солидарен с первой программисткой – у машины Бэббиджа отсутствует возможность «самостоятельно мыслить» и её аргумент правомочен, если относится к машине Бэббиджа. Тем не менее, можно предположить, что у других компьютеров могут проявиться самостоятельное мышление и творчество. Однако такие машины появились после смерти Лавлейс, т.е. её аргумент имеет силу лишь по незнанию возможностей компьютерной технологии. Аналитическая машина Бэббиджа, как и ряд других, представляемы универсальной цифровой вычислительной машиной. Если такой компьютер правильно запрограммировать, то он будет способен моделировать любую машину, в том числе и машину Бэббиджа. Таким образом, помимо возможности продуцирования оригинального вывода, даже заявленная претензия на универсальность придаёт ЦВМ креативные способности – она имитирует многообразие специальных ЦВМ, и переход от одного спецвычислителя к другому трактуется как некоторый творческий процесс (в смысле универсального креативизма).

Вся вышеприведённая аргументация, помимо логической некорректности, совершенно неприемлема для дискуссионных обсуждений, так как А. Тьюринг не представил списка требований к символьной машине, которым должна соответствовать машина, способная пройти тест Лавлейс.

Подведём итог репрезентативной парадигме креативных компьютеров. Очевидна логическая некорректность аргументации против АЛ. Иное дело – коннекционистское опровержение. Здесь А. Тьюринг основательно останавливается и предлагает концептуальный проект нейрокомпьютерной машины, который позволяет квазиэмпирически, в контексте мысленного эксперимента, отвести аргумент Лавлейс. Данное опровержение мы рассмотрели выше. В дальнейших разделах этого

исследования раскрывается нейрокомпьютерный проект (машина-ребёнок) и приводятся соображения по созданию машины-взрослого.

В заключение подраздела ещё раз подчеркнём доминирующую роль проблемы творчества в междисциплинарных исследованиях ИИ. Проекты построения креативных систем более фундаментальны, нежели чем схоластические концепции компьютеризированной реализации иллюзорного феномена сознания, так как (акцентируем!) сознание оценивается по креативным проявлениям, но не наоборот. Именно на решение проблемы творчества следует направлять философские, научные и инженерные усилия, если мы желаем достичь наискорейшего успеха в развитии технологии ИИ в частности и в развитии электронной культуры в целом посредством комплексной компьютерной реализации когнитивных феноменов. В этих исследованиях необходимой составляющей представляется нейрокомпьютеринг феноменов творчества.

Далее в работе исследуются способы компьютерной реализации творческих процессов, выдвигается тезис о приоритетности проблематики творчества перед другими направлениями философской методологии ИИ, а также теоретически обосновывается интуитивно очевидное положение о преимуществе нейрокомпьютерингового программирования креативных систем перед классическим, лингвокомпьютерным программированием.

6.2. Конструирование научной методологии: нейрокомпьютинг – базовая технология искусственного интеллекта

Выше раскрывалось положение, что именно нейрокомпьютинг обеспечивает решение доминантной проблемы ИИ – проблемы творчества. Проблема творчества, т.е. изучение способов компьютерной реализации креативных процессов, представляется доминирующей в методологических исследованиях искусственного интеллекта. Также, помимо классификации креативных систем, имеющей в большей мере теоретико-концептуальный интерес, на основе ряда частных тестов Тьюринга было показано, что, в отличие от креативного лингвокомпьютинга, креативный нейрокомпьютинг обеспечивает несравнимо большие возможности имитации продуктивной деятельности в силу коннекционистской «свободы» организации квазиалгоритмических программ продуцирования оригинальных информационных продуктов.

В данном параграфе идея применения методологии КТТ конкретизируется относительно общенаучной, теоретико-эмпирической методологии. С этой целью нейрокомпьютер представлен в формате концепта методологической парадигмы, совмещающей метафизическую и инженерную проекции на социокультурную проблематику. Доказывается тезис детерминации нейрокомпьютером материального базиса и духовной составляющей электронной культуры (э-культуры), так как нейрокомпьютинг в большей степени, нежели чем другие направления ИИ, обеспечивает решение психотехнической проблемы, суть которой – имитация или воспроизведение феноменов сознания посредством НБИКС-технологий. Доказывается, что приоритет в нейрокомпьютинге принадлежит отечественным разработкам.

Нейрокомпьютерная тематика в различных формах культуры рассматривается по-разному. В традиционной культуре нейрокомпьютер оценивается, как правило, с технических и инструментальных позиций. В электронной культуре акценты смещаются – нейрокомпьютерная тема становится значимым разделом гуманитарных и социальных наук (в частности, культурологи), изучающих факторы, тенденции и закономерности развития информационного общества. Такая роль приписывается ей в силу того, что нейрокомпьютер – исторически пер-

вое и одно из главных направлений искусственного интеллекта. Искусственный интеллект, в свою очередь трактуемый как сложная многоуровневая система междисциплинарных исследований, становится главным фактором становления высокоразвитой электронной культуры [Алексеев, 2009; Гейбатов, 2010]. В связи с этим и возрастает роль нейрокомпьютерных исследований. Из технического средства эффективного моделирования в традиционной культуре, нейрокомпьютер становится детерминантом социокультурного процесса в э-культуре. В работе аргументируется это суждение. Назовём его *тезисом нейрокомпьютерной детерминации* электронной культуры.

Проблема концептуальной совместимости терминов «нейрокомпьютер» и «культура»

Обоснование тезиса о нейрокомпьютере как детерминанте социокультурного процесса в э-культуре имеет место только в том случае, если термины «нейрокомпьютер» и «культура» в смысловом отношении представляются равнопорядковыми. Дело в том, что, на первый взгляд, «нейрокомпьютер» по статусу предмета моделирования – биологической нейронной сети – концептуально несоразмерен явлениям социальной и культурной сфер жизни общества. В условиях традиционной культуры термин «нейрокомпьютер» может употребляться в контексте более-менее удачных метафор, переносящих значения слов из нейрофизиологической области в социальную сферу. В результате этого мы имеем неисчислимое количество спекуляций, например, по поводу аналогии лево-правополушарного расщепления с культурологической дихотомией «Запад-Восток». В случае, который более приемлем для практических целей, нежели чем метафорические сравнения, это – использование нейрокомпьютерной парадигмы для осуществления эвристических приёмов изучения общества. (Данное направление применения нейрокомпьютера достаточно подробно раскрыто В.М. Розиным [Розин, 2011].)

С другой стороны, в э-культуре происходит смысловая реконструкция понятий «культура», «сознание», «интеллект» и др. В условиях данной модификации «нейрокомпьютер» не рассматривается в узком эмпирическом смысле как конкретное инженерно-техническое средство решения интеллектуальных задач, распознавания образов, кластеризации, дискриминантного анализа, многопараметрической нелинейной

оптимизации, адаптивного управления или эффективного программного параллелизма и т.п.

«Нейрокомпьютер» – это концепт методологической парадигмы, совмещающей две проекции на социокультурную проблематику:

1) *метафизическая проекция*: нейрокомпьютер как способ реализации естественной или искусственной нейронной сети – это концептуальный базис коннекционистской парадигмы сознания и интеллекта;

2) *инженерная проекция*: нейрокомпьютер – это мысленный пример, погружённый в контекст многочисленных мысленных экспериментов, наработанных за шестьдесят лет развития философии искусственного интеллекта, т.е. в контекст тестов Тьюринга (название проекции инспирировано альтернативным, «калькирующим» способом перевода статьи А. Тьюринга [Turing, 1950] как «Компьютерная инженерия и интеллект»).

Две методологические проекции – метафизическая и инженерная – собственно и обеспечивают концептуальную совместимость терминов «нейрокомпьютер» и «культура» и, тем самым, реальное вовлечение нейрокомпьютерной тематики в социокультурные проблемы информационного общества, проблемы становления и развития электронной культуры.

Роль нейрокомпьютинга проявляется в рамках нейтрально-зомбистской проблематики. Из трех стратегий приписывания ментальных характеристик компьютеру она предпочтительнее, так как зомбистские эпифеноменальные конструкции в инженерном плане совершенно абсурдны, антизомбистская программа предполагает решение сверхсложной глобальной фундаментальной задачи расшифровки нейральных кодов психической активности и искусственное воспроизведение кодовых зависимостей. Нейтрально-зомбистская позиция ориентирована на построение реально действующих проектов, и психотехническая проблема сознания решается как проблема «псевдосознания», функционально имитирующего представления о феноменах реального сознания.

Следуя нашей логике доказательства *тезиса детерминации*, раскроем роль нейрокомпьютера в ИИ. Как ранее утверждалось, термином «нейрокомпьютер» обозначается концепт методологической парадигмы, совмещающей метафизическую и инженерную проекции.

Метафизическая проекция нейрокомпьютерной тематики в электронной культуре

Нейрокомпьютерная тематика в контексте метафизических теорий сознания представляется крайне плодотворной для построения продуктивных метафор интеллектуальных систем на нейросетевой основе. К метафизическим теориям сознания принято относить дуализм, бихевиоризм, физикализм, функционализм и пр. (см. неплохой обзор Э. Вильянуэвы, переведённый на русский язык [Вильянуэва, 2006]). За нейрокомпьютерную тематику «отвечает» самостоятельная парадигма – коннекционизм (в литературе встречаются иные транскрипции – «коннектизм», «коннективизм», «коннекшизм»). Коннекционизм – это направление в философии сознания и философии искусственного интеллекта, в рамках которого предпринимаются попытки объяснить феномены сознания, а также интеллектуальные способности человека, используя концепцию нейронных сетей. Составленные из большого числа структурных единиц, каждая из которых аналогична нервной клетке определённого типа, с заданным для каждого элемента весом, определяющим силу синаптической связи с другими элементами, нейронные сети представляют собой упрощённые модели мозга. Точнее было бы сказать, что нейронная сеть – метафора человеческого мозга. «Нейрон», в свою очередь, удобная аналитическая конвенция, аналогичная «атому», «гену». Коннекционизм, как правило, ожесточённо соперничает с репрезентационизмом, теорией языка мысли и другими символично-ориентированными парадигмами сознания.

Здесь возникает два вопроса: 1) как понимается «сознание» исследователями нейрокомпьютинга; 2) где прорисовывается нейрокомпьютинг в общей панораме метафизических теорий сознания? Некоторые соображения по решению этих вопросов находятся среди метафизических теорий сознания.

Роль и место нейрокомпьютинга среди метафизических теорий сознания

Нейрокомпьютер как способ реализации естественной или искусственной нейронной сети – это концептуальный базис коннекционистской концепции интеллекта и сознания.

Роль нейрокомпьютера в метафизической проекции нашего тезиса детерминации особенно чётко выявляется в спорах между «символи-

стами» и коннекционистами. Первых считают сторонниками классического подхода к ИИ («старому доброму ИИ»). Вторых, отстаивающих приоритетную роль нейрокомпьютинга в ИИ, называют сторонниками неклассического подхода. Эти споры достаточно подробно освещены в работе Дж. Гарсона [Garson, 2002] (см. также размышления М. Кольцова по этому поводу [Кольцов, 2006-1; 2]).

«Классики» выражают точку зрения, которая доминировала тридцать лет с момента возникновения ИИ – когнитивные процессы аналогичны символьной обработке в цифровых компьютерах. В классическом рассмотрении информация располагается в виде строк символов, т.е. так же, как мы организуем данные в памяти компьютера или на листе бумаги (ленте машины Тьюринга). Процесс мышления и познания имеет сходство с обработкой информации цифровым процессором, при которой строки обрабатываются в последовательности, определяемой инструкциями программы (также организованной в символьном виде).

«Коннекционисты» (неоклассики) предполагают, что информация, как у человека, так и в компьютере хранится в несимвольном виде и определяется весовыми соотношениями узлов нейронной сети. Когнитивный процесс – динамический, гетерархический процесс, выполняющийся в нейронной сети, при котором активность каждого узла зависит от силы связей и активности соседних узлов, в соответствии с активизирующей функцией.

Коннекционистов следует разделить на три лагеря.

1. *Радикальные коннекционисты.* Они всячески стремятся уничтожить модель символьной репрезентации информации как направление когнитивной науки и философии ИИ. Классическая теория обеспечивает лишь крайне низкий уровень целостности представления данных, слабую способность к обобщению информации, распознаванию контекста и к выполнению многих других функций человеческого интеллекта. Провал классического программирования по сравнению с гибкостью и продуктивностью человеческого сознания – это симптом того, что в ИИ необходима новая парадигма.

2. *Практические коннекционисты.* Они не рассматривают нейрокомпьютинг как вызов классической теории, более того, выступают в поддержку классических представлений и ищут возможности согласо-

вания двух парадигм между собой. Несомненно, мозговая активность репрезентируется нейронной сетью. Однако имеется также символичный процессор на более высоком и абстрактном уровне описания. Для этого процессора вся необходимая информация формируется как результат функционирования нейронной сети. Например, роботы, предназначенные для работы в некотором «сообществе», «колонии» роботов, имеют два уровня организации когнитивно-компьютерных механизмов. Техническое зрение таких роботов организовано нейросетевыми средствами. Но, средствами символического процессора и, соответственно, «гибридными» механизмами связи асимболического и символического уровней представления информации, эти роботы самостоятельно должны создавать свой собственный, внутренний для колонии «язык» с целью координации «взаимодействий» в сложных условиях среды. Несомненно, более приемлем подход практических коннекционистов, вводящих способы классической обработки информации в модели нейронной сети. Они синтезируют наработки в области лингвистической философии с сублингвистическим подходом. С другой стороны, современные «символисты», усматривая узость классических репрезентативных систем знаний (продукционной, фреймовой, семантико-сетевой и формально-логической моделями), подобным образом ищут теоретических и инженерных стыковок с нейрокомпьютерными моделями ИИ. В связи с этим подход практических коннекционистов и практических символистов называется *гибридным ИИ*. Примером гибридных систем являются многоагентная методология проектирования и программирования интеллектуальных систем – основы для реализации моделей искусственного общества. Нейросетевой подход к искусственным обществам широко представлен в работах В.Л. Макарова, А.Р. Бахтизина и в деятельности лаборатории «Искусственные общества» (см. сайт <http://www.artsoc.ru>). Совершенно нелепо подходить к многоагентной методологии без учёта нейрокомпьютерной тематики.

3. *Критические коннекционисты*. Исследуя фундаментальные проблемы нейрокомпьютерной тематики (в первую очередь, что суть «нейрон»), эти исследователи коренным образом изменяют представления о материальных коррелятах человеческого сознания. Отсюда возникают и новые возможности решения психотехнической проблемы нейрокомпьютерными средствами.

Критических коннекционистов, в свою очередь, следует условно разделить на два класса – гуманитарных и технических.

1. *Гуманитарные критические коннекционисты.* Эти исследователи походят к нейрокомпьютингу с позиций социальных и гуманитарных наук. Нейрокомпьютер – это инструмент изучения внутреннего мира человека, языка, когнитивных механизмов индивида и социума. К коннекционистам данного класса следует отнести известного отечественного когнитивиста – Т.В. Черниговскую (см. тему «Язык, мозг и компьютерная метафора» в [Черниговская, 2006]). Основываясь на открытии Риззолатти и Арбиом зеркальных нейронов и зеркальных систем, в этих работах доказывается принципиальная важность имитации биологическим существом действий *другого* в нервной системе для когнитивного развития в фило- и онтогенезе. Зеркальные нейральные системы, в отличие от прозрачных (классических) систем, позволяют экономнее описать способ эволюционного возникновения языка и рефлексии как причины сознания. В свою очередь, новый взгляд на понятие «нейрон» позволяет пересмотреть нейрокомпьютерные модели и подать заявку на фундаментальную корректировку нейроматематических подходов.

2. *Технические критические коннекционисты.* Эти специалисты усматривают пробелы в существующем нейрокомпьютерном инструментарии и подходят к их исправлению с позиции естественных и технических наук. Например, А.В. Савельев изучает не свойства нейрона как такового, а диффузионные свойства нейротрансмиттинга, основываясь на моделях так называемого «биологического подхода». Данный подход заключается в детальном информационном воспроизведении тонких механизмов и закономерностей функционирования реальных нейронов. Также аргументированно показывается возможность теоретического предсказания неизвестных ранее биологических феноменов [Савельев, 2009].

Таким образом, критические коннекционисты, осуществляя ревизию категорий нейрокомпьютерной тематики, способствуют научным революциям в нейрокомпьютинге и исследованиях ИИ, по крайней мере, раскрывают необходимость смены нейральной парадигмы. В целом, оппонируя символистам, все коннекционисты составляют в настоящее

время мощный лагерь исследователей проблематики ИИ и нейрокомпьютерного решения психотехнической проблемы.

Роль нейрокомпьютера (в метафизической проекции на социокультурные проблемы) можно условно «подсчитать» следующим образом. Классический ИИ и нейрокомпьютерный ИИ получают по 1/3 наработки. Оставшаяся треть приходится на гибридный ИИ.

Инженерная проекция нейрокомпьютерной тематики в э-культуре

Нейрокомпьютер, согласно инженерной проекции, – это концепция, погружённая в контекст многочисленных мысленных экспериментов комплексного ТТ. Как ранее уже отмечалось, нейронная сеть Тьюринга претендует не более чем на параллельную машину – она реализуется средствами универсальной ЦВМ. Более близкой к нейрокомпьютерной тематике оказывается машина Корсакова, предложенная в статье этого гениального русского инженера в 1832 г. Это – за десять лет до того, когда мировая научная общественность, благодаря Аде Лавлейс, узнала об аналитической машине Ч. Бэббиджа (1842 г.) – непосредственного аналога УЦВМ, взятого почти без изменения А. Тьюрингом. Никому не известная разработка гениального отечественно инженера С.Н. Корсакова, не признанная российской императорской академией наук и опубликованная не на родном русском, но лишь на французском языке, совсем недавно была «открыта» Г.Н. Поваровым [Поваров, 2005-1]. Перевод этой статьи увидел свет только в 2007 г., а был издан в 2009 г. [Корсаков (Сыромятин), 2009].

Машина Корсакова («машина сравнения идей») состоит из пяти самостоятельных механизмов, «нейрокомпьютерная» интерпретация которых может быть следующей. Страницы (перфокарты), на которых представлены таблицы машины (страниц может быть достаточно большое количество), накладываются одна на другую. В клетках таблиц указывается наличие или отсутствие некоторых признаков (в оригинале описывались патологии различных человеческих органов). Если признак имеет место, то он отмечается прокалыванием иглой той клетки, в которой он представлен (пересечение «признака болезни» с «анатомией человеческих органов»). Вполне очевидно, что клетка может репрезентировать нейрон, а игла – синапс (в классическом понимании модели

нейрона). Каждая таблица представляет некоторую простую, обозримую идею. Ряды накладываемых друг на друга таблиц соответствуют латентным узлам «сети». Те таблицы, которые просматриваются как начальная и конечная – входные и выходные узлы. Механизмы, которые приводят в движение машину сравнения сложных идей (сложная идея – это есть ряд наложенных друг на друга таблиц), выдают результат, зависящий во многом от усилия, затрачиваемого для прокалывания страниц («весовой коэффициент»). Конечно, это не модель биологического нейрона, но нечто очень похожее (это далёкий 1832 г.!). Машина Корсакова, как мы видим, крайне разнится с машиной Тьюринга и достаточно далека от «символьного» подхода к ИИ. Несомненно, машина Корсакова обладает большим потенциалом эвристических идей относительно её применения в концептуальных построениях ИИ.

Таким образом, если машина Бэббиджа – прообраз классического компьютера, то машина Корсакова – прообраз нейрокомпьютера, своего рода протонейрокомпьютер. Отсюда следует тезис:

Машина Корсакова - протонейрокомпьютер

Удивительно, что только в последнее десятилетие отечественная и мировая общественность оживлённо заговорила о выдающемся изобретателе С.Н. Корсакове¹. По долгу службы С.Н. Корсаков обрабатывал большие объёмы статистических данных. Тягостная бессмысленность осуществления однообразных действий подвела его к идее о том, что многие рутинные операции можно производить автоматически с помощью механических средств. Ему удалось впервые в мире разработать оригинальный метод быстрой обработки больших объёмов информа-

¹ Семён Николаевич Корсаков (1787 – 1853 гг.), проживал с 1827 г. до конца дней в Дмитровском уезде, в своём имении Тарусово на берегу реки Дубны. С.Н.Корсаков был сыном талантливого военного инженера Н.И.Корсакова и племянником адмирала Н.С.Мордвинова, а его крестными были князь Г.А.Потёмкин-Таврический и княгиня Е.И.Вяземская. Вернувшись героем с войны 1812 – 1814 гг., он поступил на службу в отделение статистики Министерства юстиции и на протяжении многих лет занимался составлением статистических описаний российских городов, впоследствии возглавил отделение статистики, вышел в отставку действительным статским советником (соответствует генеральскому чину на военной службе).

ции с использованием перфокарт и создать уникальные «интеллектуальные машины», ставшие, по сути, первым шагом к современным информационным технологиям и к теоретическим основам информатики и кибернетики¹.

Таким образом, можно с полной уверенностью утверждать следующее:

Первый нейрокомпьютер разработан в России (1832 г.)

Проследим дальнейшую историю нейрокомпьютеринговой инженерии, связанную с реалиями сегодняшней компьютерной технологии. Здесь у истоков мы снова видим отечественных учёных².

Идея нейрокомпьютеринга имманентна российской ментальности, по всей видимости, по причине пространственной обширности России и разнородности населения, если придерживаться точки зрения географического детерминизма в определении факторов общественного сознания. Для демонстрации этой мысли проследим пост-корсаковскую историю нейрокомпьютеринговой инженерии, связанную с реалиями сегодняшней компьютерной технологии. Здесь у истоков мы увидим отечественных учёных.

1928 – 1935-е годы – П.К. Анохин, автор теории функциональных систем и идеи самоорганизующейся обратной связи. Работы явно или неявно используются в современном теоретическом и практическом нейрокомпьютеринге. В последующем Н. Винер значительно упростил анохинские идеи о самоорганизующейся обратной связи, обратная связь у него оказалась без «самоорганизации». Упрощение привело к тому, что кибернетика в 1990-е годы стала «соперничать» с синергетикой как наукой о самоорганизации. Хотя было достаточно и одной

¹ Научная деятельность С.Н.Корсакова была многогранна – он был также выдающимся российским учёным-гомеопатом, его открытия и разработки получили высокую оценку специалистов как в России, так и за рубежом. Свои «интеллектуальные машины» он использовал и в гомеопатии, а его уникальные методы приготовления гомеопатических лекарственных средств до сих пор широко используются в Европе с пометкой «К».

² Исторический материал любезно предоставлен редактором журнала «Нейрокомпьютеры: разработка, приложение» А.В. Савельевым.

научной дисциплины, если бы не пошли на поводу теоретической и инженерной простоты.

1930-е годы – Н.А. Бернштейн, основоположник нейробиомеханики двигательной активности. Все (!) современные роботы, так или иначе, используют эти результаты. Лишь затем появляются работы Н. Винера (1942 г.), У.Мак-Калоха и У.Питтса (1943 г.), Д.Хебба (1949 г.), Ф.Розенблатта (1958 г.). В этих работах мы видим всё большее и большее упрощение и огрубление сложнейших представлений о нейральной активности. В результате этого зарубежный нейрокомпьютинг долгое время (по сути, в 1960–80-е годы) пребывает в стагнации. В то же время отечественная школа нейрокомпьютинга достигла значительных успехов.

1960 – 1970-е годы – В.М. Глушкова, Н.М. Амосова, А.М. Касаткина, Э.М. Куслуля синтезировали цифровые М-автоматы («мыслящие» автоматы).

1981 – 1987-е годы – А.В. Савельев, разработчик методов натурального нейромоделирования, применение которых позволило сделать и предсказать ряд нейрофизиологических открытий [Савельев, 2004]. Нейрокомпьютинг в СССР институализировался.

21 апреля 1971 г. – был создан НИИ нейрокибернетики АН СССР, впоследствии НИИ НК РАН им. А.Б. Когана (г. Ростов-на-Дону). Зарубежный нейрокомпьютинг кое-как выбился из кризиса в 1986 г. благодаря «переоткрытию» метода обратного распространения ошибки (Д.И.Румельхарт, Дж.Е.Хинтон и Р.Дж.Вильямс), хотя незадолго до этого аналогичный метод активно использовался в работах А.Н. Колмогорова, С.И.Барцева и В.А.Охонина, а также А.Н. Горбаня, который принадлежал другой школе – красноярской.

К успехам отечественного нейрокомпьютинга следуют отнести такие разработки, как, например, «Эмбрион» В.Д. Цыганкова, успешно работающий в медицинских приложениях и военной технике. «Эмбрион» работал с 1960-х годов в системах наведения противоракет, основан на принципах теории функциональных систем П.К. Анохина. На западе не было и до сих пор нет ни одного подобного реально работающего аппарата, решающего задачи, которые не под силу обычным компьютерам. Следует отметить, что советские ракеты летали лучше зарубеж-

ных. Их механизм был основан на самоорганизующейся обратной связи, а не на винеровской («обратная связь минус самоорганизация»).

Кроме того, необходимо перечислить следующие проекты: нейрочипы А.И. Галушкина, используемые в управлении вертолётами, вертолётами-роботами; многопроцессорные с перестраиваемой структурой нейрочипы А.В. Каляева (г. Таганрог); нейристоры — нейроподобные матрицы А.А. Зенкина для автоматического доказательства теорем; реверберационный нейропроцессор А.В. Савельева [Савельев, 2006-2]; нейродетекторы Е.Н. Соколова (теоретическая часть) и Л.А. Шмельёва (практическая составляющая работ) по нейронам-детекторам; работы Н.П. Бехтеревой по расшифровке мозговых кодов; работы директора Института нейрокибернетики АН СССР и РАН А.Б. Когана и его коллектива; работы Н.В. Позина (1950 – 60-е годы) и его группы «Бионика» и другие работы отечественных «нейрокомпьютерщиков» [Савельев, 2004].

Живое активное состояние современных отечественных работ в области нейрокомпьютинга можно отследить по многочисленным конференциям, которые проводятся в нашей стране. Особенно следует отметить всероссийскую конференцию «Нейроинформатика», которая проводится ежегодно в МИФИ, и всероссийскую конференцию «Нейрокомпьютеры и их применение», также ежегодно проводимую на базе Московского городского психолого-педагогического университета¹. Несомненно, что работы отечественных исследователей активно использовались и используются зарубежными коллегами. Так, например, один из первых западных нейрокомпьютеров, «Адалин», разработанный Б.Уидроу и М.Хоффом (1960 г.), построен на основе адаптивного линейного сумматора (название происходит именно от данного словосочетания), впервые предложенного Я.З. Цыпкиным в 1948 г.

Если теперь рассмотреть «викиномичную» историю нейрокомпьютинга, которая представлена в многочисленных работах, выставленных в Интернете, и на основе которых написана работа А.Н. Липова,

¹ Широкий обзор нейрофизиологических моделей, используемых в методологии ИИ, представлен в [Петрунин, 2010].

раскрывающая историю нейрокомпьютинга, то мы в ней не увидим отечественных заслуг в развитии нейрокомпьютинга¹:

1942 г. – начало вики-нейрокомпьютинг; Н. Винер вместе с со-
ратниками публикует работу о кибернетике, одной из основных идей
которой является представление сложных биологических процессов
математическими моделями;

1943 г. – У.Мак-Калох и У.Питтс предлагают модель нейрона,
формализуют понятие нейронной сети в статье о логическом исчисле-
нии идей и нервной активности;

1949 г. – Д.Хебб предлагает первый алгоритм обучения персеп-
трона;

1958 г. – Ф.Розенблатт разрабатывает принципы построения пер-
септрона (математическая и компьютерная модель восприятия инфор-
мации мозгом), первого технического воплощения нейронных сетей;

1960 г. – Уидроу совместно со своим студентом Хоффом на ос-
нове дельта-правила (формулы Уидроу) разработали Адалин (ср. с
предыдущими сведениями об Адалине), который сразу начал использо-
ваться для задач предсказания и адаптивного управления;

1969 г. – М.Минский и С.Пейперт издают книгу «Perceptrons»,
где доказывают принципиальную ограниченность возможностей пер-
септронов и т.д.

Видим ли мы в данном обзоре нейрокомпьютинга сведения об
отечественных разработках? Отнюдь нет. Читать такую вики-историю
нейрокомпьютинга неприлично. Здесь полнейшее забвение отечествен-
ной истории. Аналогична вики-история и с другими именами выдаю-
щихся основателей компьютерной науки, кибернетики, синергетики и
прочих общетеоретических и частнотеоретических оснований совре-
менной технологии. В связи с тем, что «за державу обидно», предлага-
ется считать проект С.Н. Корсакова своеобразным символом возрожде-
ния отечественных идей в науке – идей, про которые сегодняшние
СМИ напрочь забыли. Задел в таком подходе к корсаковским идеям
обоснован в виде следующего тезиса:

¹ Липов А.Н. Из истории нейрокомпьютинга// Полигнозис, 2(42), 2011, С.120 – 130.

Родиной нейрокомпьютинга является Россия

Это важно для аутентичной трактовки инженерной проекции нейрокомпьютинговой тематики на сферы социокультурной жизни.

Вернёмся к мысленным экспериментам комплексного теста.

Тест Тьюринга «с мозгом»

Многие авторы частных ТТ особо подчёркивают то, что нейрокомпьютер – это наиболее адекватное средство для имитации интеллектуального и сознательного поведения. Так, например, Дж. Сёрль в знаменитом ЧТТ «Китайская комната» (1980 г.), полагал, что одно из наиболее сильных опровержений тезиса «компьютер понимать не может» происходит со стороны специалистов по «моделированию мозга». Конечно, он приводит совершенно невразумительный и, прямо скажем, глупый отвод «нейрокомпьютерного» возражения путем аналогии нейронной сети с системой водопроводных труб, в которой «соединения труб» – синапсы, а положения клапанов в трубах – «степень возбуждения сети» [Сёрль, 2006, С.17]. Многие оппоненты и сторонники Дж. Сёрля полагают, что «Китайская комната» послужила современной волне развития когнитивной науки в попытках технически опровергнуть его тезис и доказать возможность реализации (по крайней мере, имитации) феноменов сознания путём моделирования естественных когнитивных механизмов, в первую очередь, нейрокомпьютерными средствами.

Нейрокомпьютерная тематика также привлекает внимание зомбифилов (зомбистов, антизомбистов, нейтральных зомбистов) всех мастей. Так, Т. Полджер, О. Фланаган, отстаивая принцип несущественности сознания для осуществления высокоинтеллектуальной деятельности, полагают, что более сильное испытание, нежели тест Тьюринга, представляет «тест Тьюринга с мозгом» («Brainy-Turing-Test»). Этот тест в дополнение к поведенческой эквивалентности требует эквивалентность нервной системы, т.е. помимо тождественности ввода-вывода нужна тождественность с мозговой активностью [Фланаган, Полджер, 2006, С. 103 – 112].

В тесте Харнада «тест Тьюринга с мозгом» знаменует «тотальную неотличимость» компьютерной системы от человека, включая мельчайшие внутренние нюансы телесного строения. Это – уровень микрофизической (в контексте нашей работы – нанотехнологической) неотличимости естественных и искусственных сознательных систем.

Таким образом, многими исследователями ТТ нейрокомпьютер воспринимается в роли мысленного примера, обозначающего высшую ступень возможности решения психотехнической проблемы.

Со своей стороны, хочется отметить, что в литературе иногда встречается мультимедийный ТТ, способный тестировать х-системы в диапазоне симультанных способов репрезентации информации [Clifton? 2003]. Очевидно, что с видео- и аудиорядами (в мультимедиа они существенным образом превалируют над текстовым форматом представления информации) эффективно справиться лишь нейрокомпьютер. Классический ИИ к мультимедийному тесту не готов. Мультимедиа, в свою очередь, это базовый инструментальный выражения когнитивных феноменов э-культуры-1 и когниций превращенной формы традиционной культуры – э-культуры-2.

В завершение изучения метафизической и инженерной проекции нейрокомпьютерной тематики на э-культуру подведём итоги.

Значимость образной информации (за изучение её «ментальных» параметров должен отвечать мультимедийный ТТ) – 90 % от вербальных форм. Таким образом, роль нейрокомпьютера в реализации полноценного ТТ, отвечающего требованиям традиционных форм э-культуры (э-культуры-1), не говоря уже об остальных формах (2-й и 3-ей), следует оценивать как 9/10. Учитываем значение показателя роли нейрокомпьютера, полученного ранее из анализа метафизической проекции (1/3). Получаем приблизительно 62 %. Конечно, это крайне условная оценка. Тем не менее, мы имеем возможность «количественно» оценить степень участия нейрокомпьютера в развитии э-культуры.

Таким образом, на основе вышеизложенного, можно сделать следующие выводы.

1. Нейрокомпьютер представляется одним из главных детерминантов развития э-культуры, так как, будучи составной частью междисциплинарных исследований ИИ, он существенным образом конституирует её материальные и духовные составляющие.

2. В э-культуре-1, реализуя традиционные формы культуры на базе НБИКС-технологий, нейрокомпьютер выполняет задачи, характерные для сегодняшнего уровня развития технологии ИИ. Степень участия в «социокультурном» процессе мы оценили на 62 %. Значение этого условного показателя целесообразно в некоторой форме учитывать в распределении средств поддержки исследований ИИ. Остальные 38% приходятся на другие исследования ИИ.

3. В э-культуре-2 нейрокомпьютер, реализуя превращённые формы сознания, выступает в качестве необходимой компоненты систем виртуальной реальности или в качестве функционального заменителя некоторых участков коры головного мозга, например, лечение болезни Паркинсона, устранение фантомных болей – это прерогатива нейрокомпьютерных чипов.

4. В э-культуре-3 при решении психотехнической проблемы роль нейрокомпьютера становится наиболее значимой. Укажем лишь на недавний факт признания со стороны Президиума РАН фундаментальных перспектив, открывающихся во всех сферах э-культуры благодаря предполагаемой реализации проекта расшифровки нейродинамических кодов субъективной реальности (это – нейрокомпьютерная проблематика!). Проект, предложенный 40 лет назад выдающимся отечественным философом Д.И. Дубровским, лишь сегодня находит практическую поддержку в плане реализации. Доклад Д.И. Дубровского, прозвучавший на Научной сессии Общего собрания РАН 15 декабря 2009 г., имел следующее название: «Мозг: фундаментальные и прикладные проблемы. Философские подходы к проблеме «мозг и психика»: к вопросу о расшифровке нейродинамических кодов явлений субъективной реальности». В проекте нет рассуждений по поводу имитации феноменов сознания. Ставится радикальный вопрос относительно реального воспроизведения или, собственно, фактической реализации феноменов сознания на базе электронных технологий. Проект широко известен как «информационный подход к сознанию» (точнее, к субъективной реальности). Перспективы, которые он сулит, очевидны – человек приобретает способность по своей воле оперировать нейродинамическими носителями феноменов сознания, управлять энергетическим обеспечением этих операций, в том числе соответствующими биохимическими процессами; изменять программы действий, следовательно, изменять их кодовые нейродинамические структуры; расширять контуры психической регуляции; интенсифицировать творческие процессы; создавать новые ресурсы психической саморегуляции, причем не только функциональной, но и нравственной. Снова повторю – это специфически нейрокомпьютерная тематика. Проект – глобальный и, несомненно, имеет не только стратегический характер для развития э-культуры страны, но и судьбоносное значение для человеческой цивилизации в целом.

5. Несмотря на то что нейрокомпьютерный проект расшифровки кодов нейромозговых коррелятов психических феноменов поражает воображение и захватывает дух, на мой взгляд, предпочтительнее выбрать менее радикальный путь решения психотехнической проблемы. Необходимо (возможно, параллельно с проектом расшифровки нейродинамических кодов) добиваться частных тактических успехов на пути аппроксимативной функциональной имитации феноменов сознания. Здесь перспективны проекты искусственной личности и искусственного общества. В первом проекте решаются вопросы создания «псевдосознательных» механизмов имитации интеллектуальной, эмоциональной, волевой сферы, изучаются проблемы компьютерного воспроизведения морального поведения, свободы выбора, моделируется самость, «Я». Во втором проекте изучаются искусственные агенты, устанавливающие «социальные связи» на основе моделирования всего многообразия факторов, присущих реальным человеческим общностям – материально-производственных, экономических, идеологических, морально-нравственных, религиозных и др. Решение отдельных задач, возникающих при реализации проектов искусственной личности и искусственного общества, т.е. частные незначительные тактические успехи, на мой взгляд, – это и есть правильная стратегия решения психотехнической проблемы.

В этой стратегии нейрокомпьютерным когнитивным механизмам отводится базовая реализационная роль, нейрокомпьютерная технология представляется ключевой предпосылкой конструирования социокультурной реальности.

Поэтому с полной уверенностью можно сказать:

Нейрокомпьютер – основная технологическая детерминанта развития электронной культуры
--

6.3. Конструирование инженерной методологии: комплексная машина К орсакова–Т ьюринга – фундаментальный проект интеллектуального компьютера

В философских исследованиях любой проблемы главная задача состоит в нахождении фундаментальных универсалий, эпистемологических средств, моделей, метафор, концепций – всего того, на чём, собственно, базируется человеческое представление мира. В философских исследованиях ИИ главная задача заключается в выявлении фундаментальной компьютеринговой метафоры, посредством которой человек воспринимает и понимает окружающий мир – природу, отношения с другими людьми, самого себя. Сегодня в когнитивных исследованиях превалирует метафора вычислений (исчислений) когнитивных феноменов, которая выросла на почве англо-саксонской культуры. Современный компьютер – порождение того, что принято называть «западным менталитетом». Западная компьютеринговая метафора, конечно, сильна, но недостаточна для реализации когнитивных феноменов, которые рассматриваются в контексте других культур. Например, когнитологическая проблематика сознания несоизмерима для англо-американской традиции и восточной традиции трактовки слова «сознание» [Алексеев, 2010-2]. Здесь возникают проблемы так называемого радикального перевода, – носители разных языков просто не способны понять друг друга. Превосходной инженерной проекцией современной когнитивно-компьютерной метафоры представляется машина Бэббиджа – Тьюринга, для которой характерно отождествление ментальных феноменов с логическими состояниями вычислительной машины. Данная метафора породила современный когнитивный функционализм, согласно которому когниция – это функция сложной компьютерной системы, по большому счёту, не зависящая от конкретного способа реализации, – будь то человеческий мозг, тина марсианина или электронные чипы технического устройства. С когнитивным функционализмом, однако, возникли глубочайшие проблемы, инспирированные как неспособностью функционализма выразить переживания от первого лица (мои личные психические состояния), так и примитивностью компьютерной метафоры. В связи с этим ещё раз повторим вопрос Д. Деннетта: разве может современный и, однако, априори старый и дряхлый компьютер

выразить многообразие ментального [Dennett, 1984]? «Не может!» – отвечает Д. Деннетт и предлагает крайне изощрённый гетерофеноменологический метод, который в некоторой степени восполняет пробелы современного компьютеринга нейрокомпьютерной методологией. Но в его методе мы видим аналитический подход, ориентированный на лингвокомпьютинг абстрактных сущностей – мемов (это социокультурные гены). Имеются и иные подходы по восполнению ущербности существующей компьютерной метафоры. Однако все они, как правило, так или иначе оказываются вкоренёнными в метафору Бэббиджа – Тьюринга, либо, по крайней мере, принципиально выразимы средствами данной метафоры.

На наш взгляд, имеется более простой способ обогатить представление о компьютеринге когнитивных феноменов. Предлагаем воспользоваться компьютерной метафорой, предложенной 180 лет назад С.Н. Корсаковым.

История проекта С.Н. Корсакова в философско-методологических исследованиях искусственного интеллекта

Изучение забытого исторического наследия неразрывно с историей его возрождения, причем это непосредственно касается социально-эпистемологических характеристик развития философской методологии ИИ. Поэтому начнём с того, что в 2004 г. в оргкомитет Всероссийской конференции «Философия искусственного интеллекта» поступила заявка выдающегося отечественного историка науки и техники – Гелия Николаевича Поварова. Предлагался сенсационный доклад «С.Н. Корсаков – русский пионер искусственного разума» [Поваров, 2005-1]. К прискорбию, доклад не состоялся – Г.Н. Поваров скоропостижно скончался. В докладе сообщалось о том, что в начале 1990-х годов в Германии был найден архив, вывезенный из Франции, оккупированной в годы Второй мировой войны. В архиве находилась интереснейшая статья русского инженера начала 19 в., описывающая уникальный компьютер, состоящий из «машин сравнения идей». Г.Н. Поваров представил статью на английском языке [Povarov, 2001]. Благодаря этому международная общественность узнала о машинах Корсакова. Г.Н. Поваров не без основания предлагал считать Семёна Николаевича Корсакова пионером искусственного интеллекта не только в масштабе отече-

ственной, но и мировой науки. Мы также вполне правомочно выдвигаем тезис:

**С.Н. Корсаков – пионер когнитивных исследований в
масштабе мировой науки**

Дело в том, что в проекте Корсакова явно обозначена когнитивная установка: 1) когниция обозначена как «сложная мысль»; 2) механизм реализации когниции предствлен механической моделью анализа «сложной мысли», обработки «простых мыслей» и получения из них «сложной мысли»; 3) разработана инженерная конструкция, обеспечивающая машинную (компьютерную) реализацию данной модели.

Когнитивная установка проекта С.Н. Корсакова

Когнитивная установка С.Н. Корсакова состоит в следующем. Идеальные мысли следует облечь в физическую оболочку. «Материальная оболочка» мыслей по-корсаковски совершенно отличается от кругов Эйлера, диаграмм Венна и других подобных способов визуального отображения экстенсионалов (объемов понятий). Корсаковская *«материализованная мысль» – стопка перфокарт, истыканная иглами.*

К материально воплощённым «мыслям» можно применять механизмы реализации рутинных операций, которые человек осуществляет в уме при обработке сложных идей. Предлагается не одна, абсолютно на всё способная машина (как это характерно для УЦВМ А.Тьюринга), а набор из пяти машин: 1) *линейный гомеоскоп с неподвижными частями*, который выбирает из набора сложных идей, представленных в специальной таблице (перфокарте), все те, которые содержат какую-либо другую известную сложную идею со всеми ее элементами, показывая искомые идеи путем самопроизвольной остановки во время выполнения заданной операции; 2) *линейный гомеоскоп с подвижными частями*, выполняющий те же функции, что и предыдущий, и, сверх этого, способный моментально выделить и отделить те части сложной идеи, которые соответствуют (или не соответствуют) аналогичным частям других идей; 3) *планишетный гомеоскоп*, показывающий совпадение частей двух сложных идей, между которыми проводится сравнение (количество элементов может достигать миллиона); 4) *идеоскоп*, который при помощи специальной таблицы выявляет следующее: все схо-

жие элементы сложных идей, обрабатываемых в ходе текущей операции; элементы одной сложной идеи, отсутствующие в другой идее, с которой производится сравнение; элементы, которые отсутствуют в данной сложной идее, но присутствуют в той, с которой производится сравнение; элементы, отсутствующие в сравниваемых сложных идеях, но являющиеся частью других сложных идей, занесенных в таблицу. 5) *простой компаратор*, производящий те же операции, что и идеоскоп, но лишь в отношении двух сложных идей, сравниваемых между собой. Он способен обрабатывать всего несколько десятков элементов таких идей, но выгодно отличается от остальных устройств тем, что не требует специальной таблицы.

Способ работы машины прекрасно иллюстрирует *гомеопатический пример*, предложенный самим С.Н. Корсаковым. Предметом профессиональных исследований и личных пристрастий С.Н. Корсакова выступали сложные предметные области – демография и гомеопатия. Сложность предмета определяется столь большим разнообразием феноменов и связей между ними, что человеческий разум не способен сразу всех их охватить и вычленив нечто полезное из хаоса предполагаемых причинно-следственных связей. Для того чтобы удержать всевозможные сочетания идей в виде близких и дальних связей, собственно и требуются усилители человеческого интеллекта, механизмы и машины. С.Н. Корсаков предлагает механические табуляторы, которые совместно моделируют сложную логику работы человека при анализе «сложных мыслей».

Для первой предметной области С.Н. Корсаков предложил метод обработки больших объёмов статистических данных и их каталогизации, согласно которому многие рутинные операции можно производить с помощью перфокарт и механизмов. Подобный метод он предлагает и для второй предметной области – для гомеопатии. В выборе данного предмета исследований мы видим желание автора когнитивных машин достичь успехов в медико-инженерном врачевании. «Интеллектуальные машины» успешно используются в современной гомеопатии – гомеопатические средства, подобранные по корсаковской методике, принято помечать литерой «К» в честь автора. С.Н. Корсаков сформировал объёмные «базы данных» и снабдил их «программой» быстрого поиска гомеопатических средств, в зависимости от многообразных симптомов

заболевания. И успешно применял машину для лечения больных в своей округе, став, пожалуй, первым в мире врачом-инженером.

В качестве примера, который иллюстрирует действия машины, С.Н. Корсаков предлагает метод гомеопатической диагностики. В одних части таблицы он вписывает буквенные обозначения различных частей и органов, т.е. фиксирует своеобразную «топографию» человеческого тела (голова, нос, глаза, желудок и др.). В другой части таблицы, которая называется «патографией» человека, цифрами обозначаются различные недуги (опухоли, покраснения на коже, стреляющая и режущая боль, жар, озноб, обморок и т.д.). Например, если буква е обозначает грудь, а цифра 4 – режущую боль, то квадрат е4 соответствует режущей боли в груди. Для фиксации данного симптома в центр квадрата е4 втыкается игла. Подобным образом фиксируются остальные симптомы. В таблице, содержащей по сто ячеек по вертикали и по горизонтали, с помощью игл можно представить десятки тысяч различных симптомов, осуществляя тем самым сложную конструктивную диагностику. Далее машины приводятся в движение, независимо от человека определяются результаты сочетания сложных идей и осуществляется подбор медицинских рекомендаций для лечения.

Таким образом, «интеллектуальные машины» Корсакова – первый действующий когнитивистский проект, в котором четко прослеживаются составляющие когнитивной установки – модель когнитивного феномена и машина обработки этой модели. Машина Корсакова – это фундаментальная когнитивная метафора, которая, несмотря на срок давности, востребована в контексте современной культуры. Проследим историю возникновения ряда других фундаментальных метафор.

Справедливости ради отметим, что в средневековье имелся более древний проект когнитивной системы, в которой участвовали модели ментального и механизмы обработки данной модели. Имеется в виду машина испанца Раймонда Луллия (ок. 1232 – ок. 1316). Изначально машина предназначалась для юридической практики и реализовывала постглоссаторский юридико-догматический логико-символический метод толкования собраний источников римского права (глосс) применительно к конкретным житейским ситуациям, а также метод координации двух силлогизмов, один из которых представлял положения естественного права, другой – положительного права. Машина позволяла

выводить заключения из данного полисиллогизма. В последующем проект послужил основой труда «Великое искусство» (Ars magna), который претендовал на универсальный способ открытия любых истин бытия. Устройство машины состояло из подвижных концентрических бумажных кругов, разделённых поперечными линиями на отделения, на которых помечались общие понятия. Вращение кругов приводило к множеству новых комбинаций. Идея пользовалась успехом среди ученых вплоть до 18 в., хотя за абсурдность и бессмысленность комбинаторики универсалий подвергалась уничижительной критике со стороны философов «когнитивистской» ориентации (Декарта, Гегеля, Пирса) и карикатурному осмеянию в художественной литературе в образе луллитов (Дж. Свифт, «Путешествие Гулливера»). Тем не менее, здесь мы видим первую попытку *механического воспроизведения феноменов человеческого мышления*. Именно поэтому проект Р. Луллия следует считать первым проектом искусственной когнитивной системы, но в контексте современной электронно-технической базы данный проект представляется неустаревавшим, в отличие от проекта Корсакова и Бэббиджа.

Итак, корсаковский проект (1832 г.), ориентированный на демографию и медицину, начинает получать международную известность в начале 21 в. Проект был реализован, работал и успешно применялся.

Также известна аналитическая машина Бэббиджа (1833 г.), предназначенная для решения экономических задач.. Этот проект получил международное признание в 1842 г. благодаря статьям «первого программиста» Ады Лавлейс, переведённым итальянцем Л. Менабреа с английского языка на французский, который был в те годы международным (Как мы ранее отметили, С.Н. Корсаков представил свой проект также на французском языке.) Проект не был реализован.

Еще один проект – электромеханических табуляторов американца Голлерита, использующих перфокарты, – появляется в 80-е годы 19 в. Служил для переписи народонаселения Америки в 1880 – 90 гг. Фирма IBM «выросла» на табуляторах Голлерита, в Московском политехническом музее имеется соответствующая экспозиция. Проект действующий.

Аналогом машины Бэббиджа стала машина Тьюринга (1936, 1950 гг.), в которой использовалось формальное определение понятия

алгоритм [Turing, 1936, 1950]. Проект действующий, служит основанием современной компьютерной технологии.

Машина Корсакова в комплексном тесте Тьюринга

Для демонстрации роли и места машины Корсакова в КТТ рассмотрим её в формате системы, которая выступает в качестве средства реализации комплексного ТТ. В табл.1 мы видим, что КТТ в составе представленной совокупности тестов – это очередной частный ТТ, выполняющий особую задачу (само-)собираения частных ТТ с целью аппроксимативно полного покрытия многообразия когнитивных феноменов и решения общих задач их атрибуции, предикации и реализации.

Машина Корсакова – это подсистема инструментария, реализующего КТТ. Другой подсистемой является машина Тьюринга – Бэббиджа. Такое совместное применение двух фундаментальных метафор обусловлено не только тем, что машина Корсакова – это коннекционистская метафора вычислений, а машина Тьюринга – репрезентативная метафора. Мы решили совместить две метафоры, так как в машине Тьюринга налицо возможность машины автоматически сохранять состояния и переходить их одного состояния в другое. В машине Корсакова такого нет. С другой стороны, у Тьюринга нет того, что имеется в машине Корсакова – особого способа представления гетерономной информации, что особенно четко проявляется при полисемантической интерпретации лингвистических средств инструментария КТТ.

Поскольку, согласно нашим представлениям, комплексный тест Тьюринга – это частный ТТ, выполняющий специфические функции комплексирования других частных тестов, включая самого себя, т.е. вследствие синергетической специфики, инструментарий КТТ предполагает *гетерархическую* организацию и гибридную (лингво- и нейрокомпьютерную) программную реализацию. Общие идеи гетерархических систем задают систематичность, рекурсивность, гетерогенность и динамичность инструментальных средств ЧТТ, инкорпорируемых в КТТ (Янковская, 2008; 2011). Систематичность предполагает виртуальную системную целостность частных тестов. Рекурсивность выражается автогенеративным программированием инструментальных средств. Гетерогенность проявляется в функционировании в рамках единой «саморазличающейся» системы взаимозависимых уровней, каждый из которых обладает собственной логикой развития. Динамичность предпо-

лагает изменение отдельных уровней и реструктурирование системы в целом под влиянием других уровней при сохранении концептуального единства, обозначенного идеей комплексного теста Тьюринга.

Лингвистические средства машины Корсакова в контексте КТТ обеспечивают библиографическую работу идентификации, систематизации, унификации, координации, обобщения, дифференциации, интеграции ЧТТ и их компонентов. Синтаксически они выразимы нестрогим первопорядковым исчислением предикатов, интерпретированном на 3D-онтологии, объединяющей суждения трех общих планов, специфицируемых в соответствии с последующей детализацией: феноменологического (от первого лица), научно-теоретического (биологических, нейрофизиологических, психологических и других дисциплин), инженерно-компьютерного (проекты реализации ЧТТ). Данные три общие плана при условии последующей конкретизации предполагают полионтическую интерпретацию, либо, если назвать по-иному, полимерную семантику языка комплексирования ЧТТ.

Полимерная семантика означает наличие многообразия локальных целостных подсистем интерпретации многочисленных комбинаций суждений, которые выражают следующее: 1) личные представления о психических переживаниях и явлениях социальной жизни; 2) теоретические осмысления этих явлений; 3) инженерно-компьютерные проекты. Для реализации полимерной семантики собственно и востребована машина Корсакова. В отличие от оригинального варианта, где мы видели 2D-измерения («топография» и «патография»), для нашего случая имеем три общих измерения: 1) суждения от первого лица; 2) теоретические суждения нейрофизиологического, социологического, экономического и прочих планов; 3) конкретные компьютерные проекты, взятые как в целом, так и разобранные по подсистемам.

Будем теперь «прокалывать корсаковскими иглами» 3D-перфокарты, представляющие обозначенные выше измерения. Если теперь накладывать 3D-перфокарты одна на другую и варьировать глубину проникновения игл в ряды наложенных друг на друга таблиц, то мы будем получать комбинаторные комбинации локальных целостностей, которые характеризуют, скажем, для феномена «переживание» триединство «ощущения переживания» / «теории переживания» / «про-

екты реализации переживания». И, собственно, здесь, на этапе полимерного конструирования предметной области, проявляется преимущество машины Корсакова перед машиной Тьюринга. И в первом и во втором случае мы видим произвол наблюдателя (судьи, разработчика). В оригинальном ТТ наблюдатель произвольно приписывает когнитивные феномены некоторой *x*-системе, которую он оценивает. В принципе, наблюдатель пассивен. В случае же с машиной Корсакова наблюдатель выступает в роли активного лица – он самостоятельно конструирует реальность, выступая в роли творца когнитивных сценариев «жизни», которую «проживает» в формах электронной культуры или предлагает «проживать» другим членам своего сообщества.

**Корсаковский наблюдатель «прорубает окна» в
компьютеризованную реальность,
тьюринговый наблюдатель – общается с миром
сквозь открытое окно**

Информационные средства машины Корсакова, помимо стандартных, включают в себя базы «смыслов», которые репрезентируют эпистемические генерализации «данных» и «знаний», характеризуя тем самым концентрированные «знания о знаниях». В современной методологии компьютерного моделирования «смысла» предложено большое количество интересных моделей «смысла» – психологических, лингвистических, кибернетических, синергетических, риторических, поэтических, семиотических, герменевтических, математических и др. [Алексеев, 1998; 2005-8]. В них прослеживаются три взаимосвязанные трактовки понятия «смысл».

1. *Интенциональное определение*: «смысл» – это целевая направленность, ценностная ориентированность. К данной трактовке относятся выражения «смысл жизни», «смысл истории», «смысл бытия», т.е. «делать что-то с мыслью».

2. *Контекстуальное определение*: «смысл» – это выражаемый знаком способ задания значения, т.е. более широкий контекст для значений языковых выражений, связанных текстов, образов сознания, ментальных состояний и др. Обыденное словоупотребление, соответствующее этой позиции, связывает «смысл» с идеей, сущностью, точкой зрения («в смысле того или иного»).

3. *Контентуальное определение* (от слова «контент» – содержание). Ему отвечает термин *со-мыслие*. Смысл как контекст (смысл – «вне» всякой мысли) и смысл как контент (смысл «состоит» из мыслей) – это принципиально разные уровни анализа. Контентуально трактуемый «смысл» как личностно-значимое конкретно-целостное систематическое единство «мыслей» характеризует их эмерджентность и синергичность. Имеется ряд авторитетных утверждений, что «со-мыслие» присуще специфически русскоязычному варианту употребления слова «смысл».

На наш взгляд, подход к компьютерному моделированию «смысла», который получается в результате интерпретации машины Корсакова, уверенно покрывает все эти подходы. Во-первых, интенциональное определение «смысла» фиксируется в «силе» прокалывания иглой стопки перфокарт. Во-вторых, контекстуальное определение четко прослеживается, если перфокарту Корсакова интерпретировать как контекст, в котором разворачиваются «тексты», помечающие, скажем, индикаторы информационных ресурсов в местах разметки перфокарт иглами. В-третьих, контентуальная парадигма проявляется, если учесть обозначенное выше полионтическое определение предметной области и воспринимать помеченную стопку перфокарт как единое целое. Для этого «единства» возможны не только эпистемические обобщения, но и, скажем, эмоционально-художественные представления, если соответствующим образом раскрасить индикаторы информационных ресурсов и их динамические связи. Подобного рода демонстрация модели «смысла» по своим эстетическим особенностям, на наш взгляд, сродни фрактальной красоте динамических систем.

Информационные средства машины Корсакова – это преимущественно базы «смыслов»

Программные средства машины Корсакова представляются в общем случае набором клуджей – т.е. программ, которые работают не для всех учтенных случаев спецификации и документации алгоритмов (все эти случаи не подлежат скрупулёзному учету). Каждый программный клудж отвечает за реализацию частного теста или его компоненты. В целом, мы не видим здесь никакой интегрально-систематизирующей методологии, аналогичной современным крупным программным архи-

тектурным решениям. Хотя они не исключаются. Идеал организации способа программирования машины Корсакова – свободное программное обеспечение. Хотелось бы обозначить данное направление программирования как «гетерономное» программирование, отвечающее перечисленным выше параметрам гетерархических систем.

Парадигма программирования машины Корсакова – гетерономное (клюджевое) программирование

Организационные средства машины Корсакова предпочтительно реализовать по схеме «мета-веб-сообщество». Они призваны объединить в социальную сеть специалистов различного профиля и обеспечивать междисциплинарную работу интернационального сообщества для систематического изучения способов компьютерной реализации когнитивных феноменов.

Гибридная репрезентативно-коннекционистская интерпретация машины Корсакова

Сегодня предложены две равноправные трактовки корсаковского проекта, которые с учётом терминологических особенностей методологии ИИ правомочно назвать репрезентативной (символьной) и коннекционистской (нейральной) интерпретациями машины Корсакова. *Репрезентативная интерпретация* была предложена А.С. Михайловым в работе [Михайлов, 2009], где детально представлялось теоретико-множественное описание способа функционирования всех пяти машин. Ярко показано, что «интеллектуальные машины» Корсакова реализуют функции экспертных систем, обеспечивают эффективный многокритериальный поиск информации, фасетную классификацию, программное управление «базой знаний» и прочие функции современных систем репрезентативного ИИ.

Коннекционистская интерпретация предложена в ряде работ автора и представлена выше. Перспективы применения машины Корсакова колоссальны. Дело в том, что изменение компьютерной метафоры, которая лежит в основании когнитивной науки, автоматически влечёт за собой коренную трансформацию методологии науки, её теоретических постулатов и практических перспектив. Поэтому автор предлагает дополнить традиционную (параллельно) линейно-символьную компьютерную метафору Бэббиджа – Тьюринга – Неймана более мощной, объёмно-

нейросетевой компьютерной метафорой Корсакова. Технологическую модернизацию предлагается осуществлять сквозным способом, от компьютеринговой системотехники элементарных логических вентилей до 3D-онтологии когнитивных феноменов.

В эпоху э-культуры следует говорить о явной технологической детерминации социокультурного процесса. Современная базовая концептуальная детерминанта э-культуры – это метафора вычислимости когниций в соответствии с компьютерной парадигмой Тьюринга. Как мы показывали на протяжении данной работы, машина Тьюринга слишком груба для представления многообразия социокультурных феноменов. Данные феномены не вычислимы в смысле компьютеринговой дефиниции Тьюринга. Однако они могут быть вычислимы в смысле Корсакова. Для доказательства данного утверждения требуется пересмотр фундаментальных оснований теории алгоритмов, к чему, собственно, и призывает автор данной статьи. Будет модификация теории алгоритмов, возникнут новые элементарные логико-цифровые схемы – электронная культура засверкает новыми, более жизнеутверждающими оттенками. По крайней мере, современная э-культура, из-за чрезмерного отвлечения человека на технологические формы обживания социальной реальности, отдаляет человека от других людей и от самого себя. Возможно, перспективный инструментарий э-культуры, в котором проявится роль машины Корсакова как фундаментальной когнитивной метафоры, заставляющая «пробивать» все новые и новые «окна в реальность», предоставит нам больше возможностей собственно человеческого развития, не связанного жесткими алгоритмами категоризации действительности.

Конечно, здесь мы фиксируем только призывы. Однако обрисованные нами контуры новой парадигмы вычислимости, на мой взгляд, достаточно конструктивны. Автор обязуется разработать конкретный проект когнитивной машины, в которой алгоритм работы математика (машина Тьюринга) будет поддерживать квазиалгоритм работы художника (машина Корсакова), и, наоборот.

**Художник и математик совместно участвуют в развитии
электронной культуры**

В 2012 г. исполнилось 225 лет со дня рождения С.Н.Корсакова и 180 лет со дня написания эпохальной для когнитивной науки статьи. Изобретения Корсакова были столь революционны, что его современники не смогли по достоинству их оценить, как в силу субъективных причин – академической невосприимчивостью к новаторству изобретения, так и объективных причин – для практической пригодности машины Корсакова требуется мощь перспективной электронной техники. Следует должным образом использовать наследие нашего выдающегося соотечественника, более полутора века пребывавшее в забвении и адресованное даже не сегодняшнему, но, пожалуй, будущему поколению. Несомненно, актуальность возрождения проекта С.Н. Корсакова обусловлена историографическими соображениями, это крайне важно для сохранения культурного наследия России и формирования национального самосознания в эпоху массовой подмены оригинальных отечественных идей зарубежными аналогами, а порой и откровенного «экономического» обмана. В условиях стремительного развития когнитивной науки и когнитивно-технологии, становления социальных институтов когнитивных исследований, возникновения новых образовательных специальностей и учебных курсов когнитивной направленности особую важность приобретают вопросы исторических оснований когнитивистики. Изучение её истории позволяет чётче позиционировать предмет исследований, определить специфику и отличие от родственных научно-технологических направлений, погрузить исследования в социокультурный контекст, изучить перспективы развития, которые тем дальше и яснее в будущем, чем глубже и чётче просматриваются их основания в прошлом. В данной работе представилась возможность обратиться к незаслуженно забытому наследию С.Н. Корсакова и интерпретировать его работы как основы когнитивной науки. Предлагается отсчёт электронной культуры начинать не только от англо-американских, но и от российских изобретений. Появление мощных когнитивных моделей, разработчиками которых являлись отечественные учёные и инженеры, – это объективные патриотические основания для гордости по поводу вклада России в становление и развитие самой высокой технологии современности.

Тесная связь неизвестных идей Корсакова с известными идеями Тьюринга, вошедшими в стандарт любого компьютерного «юзера», сулит человеку большую степень свободы в обществе э-культуры. В этом автор усматривает *гуманистический статус* своей работы.

Ещё одна знаменательная дата – 23 июня 2012 г. В этот день мировая общественность праздновала знаменательный юбилей – 100 лет со дня рождения выдающегося английского ученого Алана Мэтисона Тьюринга. Он написал мало теоретических статей по логике, математике и вычислительной технике. Объём их невелик, однако представленные в этих работах категориально простые и потому гениальные идеи нашли широчайшее применение и многообразные интерпретации буквально во всех сферах жизни современного общества, инфраструктура которого настолько компьютеризована, что сам факт его существования и самодетерминации задаётся компьютерным инструментарием.

Этих двух гениальных инженеров разделяет столетие. Объединение их идей позволяет реформировать принципы современного компьютеринга и, по мнению автора, будет способствовать развитию человеческой цивилизации. Однако для разработки этого инструментария нужно многое сделать.

ЗАКЛЮЧЕНИЕ

РОЛЬ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА В МЕТОДОЛОГИИ ИСКУССТВЕННЫХ ОБЩЕСТВ

Современные междисциплинарные исследования искусственного интеллекта охватывают широкий спектр когнитивно-компьютерных феноменов. Философские, научные, инженерные исследования ИИ в последнее десятилетие достаточно четко дифференцированы проектами конвергентного развития комплекса НБИКС (нано-, био-, когни-, инфо-, социо-технологий). Развитию биотехнологий способствуют исследования «искусственной жизни» (artificial life), ориентированные на компьютеринг витальных феноменов. Когнитивная технология базируется на метафорах, подходах, концепциях, моделях «искусственного сознания» (machine consciousness, robot mind) и «искусственной личности» (artificial personality, computer person), соответствующих компьютерингу ментальных феноменов (квалиа, восприятие, представление, мышление, понимание, интуиция, доступность, проективность) и персональных феноменов (самость, личность, индивидуальность, моральная вменяемость, свобода, ответственность). Социокультурные явления в последнее десятилетие активно изучаются в «искусственном обществе» (artificial society) – чрезвычайно перспективной инструментрии информационных и социальных технологий, ориентированном на компьютеринг самых различных явлений общественной жизни – от материально-производственной сферы до религиозных верований. «Искусственный мир» (artificial world, synthetic world) в ряде футурологически окрашенных концепций представителей «цифровой философии» связывается с достижениями нанотехнологии, однако даже в презумпции редукционистского физикализма имеются методологически корректные идеи творить «искусственный мир» посредством интеграции достижений НБИКС, причём необходимо добавить к этому комплексу космическую технологию.

Подобного рода меганаучным, сверх-интеллектуально-ёмким проектам отнюдь не по историко-научным соображениям присваивает-

ся ярлык «исследования искусственного интеллекта». Во-первых, им имманентен сложный интеллектуальный компьютеринг, в основе которого – схемы *творческой* деятельности человека. Во-вторых, и, пожалуй, наиболее значимо то, что фундаментальные способы исследования и реализации когнитивно-компьютерных феноменов НБИКС-технологий инспирированы методологическими и мировоззренческими идеями, концепциями, моделями, методами, средствами, наработанными за шестьдесят лет в философских исследованиях искусственного интеллекта. В связи с этим, скажем, «философия когнитивной науки», «философия высоких технологий» или так называемая «философия постчеловечества» – лишь некоторые вариации философии ИИ. Именно в философии ИИ впервые была обозначена идея компьютеринговой реализации самого существенного для *homo sapiens* когнитивного феномена – *интеллекта*. Все прочие когнитивные феномены – витальные, психические, социальные – производны от интеллекта как целостной интегративной способности человека упорядочивать чувственные данные, рационально рассуждать и нормативно регулировать активность.

Проект «искусственного общества»¹ представляется наиболее значимым инструментарием в ряду других возможных способов анализа состояния, планирования и прогнозирования путей общественного развития. В самой основной части этого инструментария – в методологической – особенно важными представляются следующие собственные и производные идеи, связанные с именем А.М. Тьюринга: 1) машина Тьюринга как компьютерная дефиниция «алгоритма» (1936 г.); 2) функционализм машины Тьюринга (предложен Х. Патнэмом в 1960 г.) как компьютерная дефиниция «сознания», и вытекающая из данного определения функционалистская психофизическая парадигма, изначально решающая проблему «дух/тело» путём отождествления психи-

¹ Термин «искусственно общество» крайне полисемичен и его использование имеет давнюю историю. В рамках наличных у автора электронных ресурсов JSTOR и Wiley, первое употребление слова «artificial society» обнаруживается в обзорной работе 1863 г., описывающей способы конструирования социального поведения горилл и тропических негров [Owen, 1863, P.52].

ческих феноменов с логическими состояниями машины Тьюринга и доминирующая в современной философии сознания в различных вариациях; 3) тест Тьюринга как компьютерная дефиниция «интеллекта» и формальное определение понятия «искусственный интеллект» (1950 г.); 4) разнообразные версии и интерпретации теста Тьюринга (с 1950-х годов по сегодняшний день) как компьютерные дефиниции многообразных физических, витальных, ментальных, персональных и социальных феноменов, таких как «жизнь», «творчество», «познание», «понимание», «любовь», «свобода», «общество» и пр., предложенные в работах Дж. Маккарти, Д. Деннетта, Н. Блока, К. Колби, Дж. Серля, Р. Френча, Ю. Геновой, С. Ватта, А. Сломана, С. Брингсйорда, Дж. Лукаса, Р. Пенроуза, Дж. Мура, Дж. Баресса, С. Харнада, П. Швайзера, Дж. Поллока, Р. Кирка, Д. Чалмерса, В.К. Финна и др.

В данной работе предлагался комплексный тест Тьюринга, который в контексте этого перспективного проекта следует рассматривать как компьютерную дефиницию «общества» и формальное определение понятия «искусственное общество». По мнению автора, именно интегральное использование частных компьютерных дефиниций, которые предлагаются в различных версиях тестов Тьюринга, способствует динамичному охвату полидисциплинарной проблематики искусственных обществ – самой сложной проблематики на сегодняшний день в науке, культуре, технологии и наиболее перспективной для формирования общих и частных стратегий общественного развития.

В системе исследований искусственного интеллекта методология искусственных обществ (ИО) обладает фундаментальным и интегральным статусом. *Фундаментальная роль* приписывается в силу общих положений о роли и месте научной методологии в структуре знания современного человека. В исследованиях ИИ, в контексте которых рассматривается проблематика ИО, выделяются следующие уровни [Алексеев, 2006]: 1) *философский*, который формируется рефлексией над а) мировоззренческими ориентирами современного человека, связанные с компьютеринговой модификацией традиционных представлений о смысле жизни, свободе, социальном идеале; б) методологическими способа-

ми организации междисциплинарных исследований когнитивно-компьютерных феноменов; 2) *научный*, на котором а) выдвигаются теоретические конструкции (идеи, модели, системы) и б) исследуются прототипы когнитивно-компьютерных систем – эмпирические модели, которые верифицируют и эвристически обогащают теоретические положения; 3) *практический уровень*, на котором а) проектируются, внедряются, развиваются инженерные системы и технологии и б) используются компьютерные технологии в быту и повседневности. Система исследований ИИ целостна, так как на бытовом уровне действует конкретный индивид – носитель мировоззрения и поэтому уровень 3.б замыкается на уровень 1.а.

Методология искусственных обществ, формируя категориальные способы обживания современным человеком мира электронной культуры, цементирует все выделенные уровни. Фундируя общие параметры и специфические характеристики в философских, научных и практических методах, методология ИО призвана обеспечить создание адекватного концептуального инструментария как «прибора добывания знания» (В.Л. Макаров), «культурной универсалии» (В.С. Стёпин), «аппарата отображения» (К. Лоренц), «воспринимающего аппарата» (К. Поппер), «априорной конструкции» (И. Кант). Используя такой когнитивно-компьютерный путеводитель по просторам информационного общества, человек рационально конституирует самого себя и конструирует собственное социокультурное окружение. Сегодня о полномесном инструментарии ИО можно судить лишь в сослагательном наклонении, однако автор не видит особых принципиальных трудностей в его построении.

Интегральная роль, приписываемая методологии ИО в исследованиях ИИ, объясняется предметно-содержательными тенденциями комплексного компьютеринга всех сфер жизни человека и общества, начиная от биологических проявлений и завершая трансцендентными феноменами религиозных верований. Функция интеграции различных направлений ИИ реализуется комплексным тестом Тьюринга, который обеспечивает конкретизацию, как правило, абстрактных когнитивно-

компьютерных моделей под нужды субъекта – пользователя инструментарием ИО.

В методологических исследованиях ИО одна из главных задач состоит в выявлении фундаментальных компьютерно-ориентированных культурных универсалий, эпистемологических средств, моделей, метафор, концепций – всего того, на чём, собственно, базируется человеческое представление мира. Главная заслуга А.М. Тьюринга состоит в том, что он предлагает фундаментальные концепции компьютеринга в форме машины Тьюринга и теста Тьюринга. В данной работе эти идеи были продолжены и, по мнению автора, была охвачена более обширная сфера индивидуальной и общественной жизни, нежели чем «интеллект», для тестирования которого предназначался оригинальный тест.

Комплексный тест Тьюринга – это когнитивно-компьютерная дефиниция общества электронной культуры

«Искусственное общество» - лингвонейрокомпьютерная реализация комплексного теста Тьюринга

Комплексный тест Тьюринга, таким образом, не является «тестом тестов Тьюринга».

Статус теста вполне прагматичен.

БИБЛИОГРАФИЯ

- [Алексеев, 1994-1] Алексеев, А.Ю. Персоналистическая концепция информатизации человеко-технических систем [Текст] // А.Ю. Алексеев, В.В. Деев // Информатика и науковедение. Труды 3 Междунар. науч. конференции 28-29 нояб. 1994 г. – Тамбов: МИНЦ - 1994 г., с.3-4
- [Алексеев, 1998-1] Алексеев, А.Ю. Гуманизм, персонализм и информатика (к общетеоретическим основам моделирования искусственной личности) [Текст] / А.Ю.Алексеев // Здравый смысл, 1998. № 3/7, С. 52-56
- [Алексеев, 1998-2] Алексеев, А.Ю. К осмыслению модели смысла [Текст] / А.Ю.Алексеев // Здравый смысл, № 7, 1998, С. 48-52
- [Алексеев, 1998-3] Алексеев, А.Ю. Искусственная личность. [Текст] / А.Ю. Алексеев // Труды конференции «Междисциплинарный подход к изучению человека», М.: Институт микроэкономики, 1998, С.28-31
- [Алексеев, 1998-4] Алексеев, А.Ю. Критерии целостности гуманистической информационной технологии. [Текст] / А.Ю. Алексеев // Труды Международной московской конференции гуманистов «Наука и здравый смысл в России. Кризис или новые возможности». – М.: 1998, С. 202-207
- [Алексеев, 1998-5] Алексеев, А.Ю. Программа информационно-технологического и организационно-структурного моделирования человека. [Текст] / А.Ю. Алексеев, В.В. Деев, В.В. Насыпный, В.В. Зивенко, А.Ю. Плюснин, А.С. Абаджан // Здравый смысл, № 7, 1998, С.45-48
- [Алексеев, 2001] Алексеев, А.Ю. Паранормализация искусственного интеллекта. [Текст] / А.Ю. Алексеев // Поругание разума: экспансия шарлатанства и паранормальных верований в российскую культуру XXI века / Тезисы к международному симпозиуму «Наука, антинаука и паранормальные верования. Москва, 3-7 октября 2001 г. М: «Российское гуманистическое общество», 2001 г., С.8-11
- [Алексеев, 2002-1] Алексеев, А.Ю. Глобальные реальности XXI века. От переводчика. К статье П. Курца «Информедийная революция и перспективы глобального гуманизма» [Текст] / А.Ю. Алексеев // «Здравый смысл», № 3(24), лето 2002, М: «Российское гуманистическое общество», С.3
- [Алексеев, 2002-2] Алексеев, А.Ю. Национальные особенности логико-смысловых конфигураций // [Текст] / А.Ю. Алексеев // Приладожье и русский Север: история, традиции, современность. Материалы научно-практической конференции в честь 300-летия г.Лодейное Поле, 8 – 12 июля 2002 г. / Научн. ред. Ю.Б. Сенчихина. Сост. А.Ю.Алексеев и С.Ю.Карпук. – М.: РГО, 2003 – 208 с. – С. 12-17
- [Алексеев, 2004] Алексеев, А.Ю. Компьютерное моделирование смысла (философско-антропологический анализ). Автореферат диссертации на соискание ученой степени кандидата философских наук, М.: МГУ им. М.В. Ломоносова, 2004 г.
- [Алексеев, 2005-1] Алексеев, А.Ю. К проблеме приписывания ментальных свойств компьютеру [Текст] / А.Ю. Алексеев, А.В. Бондарь, Н.С. Скрипшевская, С.А. Игнатов, Н.В. Карпук // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛТ, 2005. – 280 с. – С. 88-92

- [Алексеев, 2005-2] Алексеев, А.Ю. Социокультурные ориентиры компьютерных моделей «смысла» [Текст] / А.Ю. Алексеев, А.Л. Дрозд, А. Кислюк, Н.В. Маклашевская, А.В. Скрябин // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С. 116-120
- [Алексеев, 2005-3] Алексеев, А.Ю. Проблема зомби и перспективы проектов искусственной личности и искусственного общества [Текст] / А.Ю. Алексеев, Т.А. Кураева, А.К. Тумасян // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С. 26-33
- [Алексеев 2005-4] Алексеев, А.Ю. Философия искусственного интеллекта [Текст] / А.Ю. Алексеев, Н.М. Смирнова // Журнал «Вопросы философии», № 9, 2005 - С. 68-80
- [Алексеев, 2005-5] Алексеев, А.Ю. К проблеме приписывания ментальных свойств системам [Текст] / А.Ю. Алексеев, А.В. Гибулянова, Н.В. Карпук, А.К. Тумасян, М. Шпёррле // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. Под ред. В.А. Лекторского и Д.И. Дубровского – М.: ИФ РАН, 2005. – 400 с. – С. 21 – 24
- [Алексеев, 2005-6] Алексеев А.Ю., Дрозд А.Л., Кислюк А., Маклашевская Н.В., Скрябин А.В. 2005. Социокультурные ориентиры компьютерных моделей «смысла» //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.116-120
- [Алексеев, 2005-7] Алексеев, А.Ю. Уровни изучения искусственного интеллекта [Текст] / А.Ю. Алексеев // Методологические и теоретические аспекты искусственного интеллекта. – М.: МИЭМ, 2005. – 192 с. – С. 23-35
- [Алексеев, 2005-8] Алексеев, А.Ю. Мировоззренческо-методологическая функция философии искусственного интеллекта [Текст] / А.Ю. Алексеев // Преподавание философии в вузе: проблемы, цели, тенденции: Сб. статей Всероссийской научно-методической конференции / Под ред. А.М. Арзамасцева. – Магнитогорск: МГТУ, 2005. – 320 с. – С. 187 – 193
- [Алексеев, 2005-9] Алексеев А.Ю., Кураева Т.А. 2005. Проблема зомби и перспективы проекта искусственной личности //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.5-10
- [Алексеев, 2005-10] Алексеев, А.Ю. О конференции «Философия искусственного интеллекта» [Текст] / А.Ю. Алексеев // Вестник Российского философского общества, 1(33), 2005 - С. 105- 109
- [Алексеев, 2005-12] Алексеев А.Ю., Кураева Т.А., Тумасян А.К. Проблема зомби и перспективы проектов искусственной личности и искусственного общества //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.26-33
- [Алексеев, 2006-1] Алексеев, А.Ю. Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга? [Текст] / А.Ю. Алексеев // Искусственный

- интеллект: Междисциплинарный подход / под ред. Д.И. Дубровского и В.А. Лекторского. – М.: ИИИнтелЛ, 2006. – 448 с. – С. 223-243
- [Алексеев, 2006-2] Алексеев, А.Ю. Методологические и теоретические аспекты искусственного интеллекта [Текст] / А.Ю. Алексеев – [ред.] – М.: МИЭМ, 2006. – 192 с.
- [Алексеев, 2006-3] Алексеев, А.Ю. Проблемы искусственной личности: развитие междисциплинарного подхода к искусственному интеллекту [Текст] / А.Ю. Алексеев // Труды XXV межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем. Ч 2. – Серпухов: Серпуховский ВИ РВ, 2006 г. – 303 с. – С. 255-260
- [Алексеев, 2006-4] Алексеев, А.Ю. Тест Тьюринга. Роботы. Зомби. Пер. с англ. [Текст] / А.Ю. Алексеев – [ред.] - М.: МИЭМ, 2006. – 112 с.
- [Алексеев 2006-5] Алексеев, А.Ю. Роль Теста Тьюринга в методологии искусственного интеллекта // Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 апр 2006 г., г.Москва: ЦЭМИ РАН / НСМЙИ РАН, 2006 г.. URL : <http://www.scmi.ru/seminar/1-11>
- [Алексеев, 2007] Алексеев, А.Ю. Когнитивно-компьютерные модели сознания и тест Тьюринга [Текст] / А.Ю. Алексеев // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн.ун-та, 2007 г. – 210 с., С. 40-53
- [Алексеев, 2008-1] Алексеев, А.Ю. Методологические проблемы проекта искусственной личности // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн.ун-та, 2008 г. – 228 с., С. 50-74
- [Алексеев, 2008-2] Алексеев, А.Ю. Определение философских зомби [Текст] / А.Ю. Алексеев // Философские науки, № 2, 2008, С. 126-149
- [Алексеев, 2008-3] Алексеев, А.Ю. Проблема другого в компьютерных коммуникациях [Текст] / А.Ю. Алексеев // Философские науки, № 6, 2008, С. 114-136
- [Алексеев, 2008-4] Алексеев, А.Ю. Трудности и проблемы проекта искусственной личности [Текст] / А.Ю. Алексеев // Журнал «Полигнозис», № 1, 2008 г., С. 20-44
- [Алексеев, 2009-1] Алексеев, А.Ю. Компьютер оценивает человека? Комплексный тест Тьюринга против комплексного компьютерного тестирования в вузовском образовании [Текст] / А.Ю. Алексеев // Полигнозис, 4(37), 2009, С.129-141
- [Алексеев, 2009-2] Алексеев, А.Ю. Электронную культуру — в жизнь! [Текст] / А.Ю. Алексеев // Полигнозис, 4(37), 2009, С.150-151
- [Алексеев, 2009-3] Алексеев, А.Ю. Понятие зомби и проблемы сознания [Текст] / А.Ю. Алексеев // Проблемы сознания в философии и науке / под ред. проф. Д.И. Дубровского. – М.: «Канон+» РООИ «Реабилитация», 2009. – 472 с., С.195-214
- [Алексеев, 2009-4] Алексеев, А.Ю. Электронная культура: трансляция в социокультурной и образовательной среде [Текст] / А.Ю. Алексеева, С.Ю. Карпук – [ред.] – М.: Изд-во МГУКИ, 2009.–260с.

- [Алексеев, 2009-5] Алексеев, А.Ю. Исследования искусственного интеллекта – главный фактор развития электронной культуры [Текст] / А.Ю. Алексеев // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С.30-36
- [Алексеев, 2009-6] Алексеев, А.Ю. Российский студент в «Китайской комнате». К вопросу о тестировании в системе образования [Текст] / А.Ю. Алексеев // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С.152-157
- [Алексеев, 2009-7] Алексеев, А.Ю. Методологические контуры науки об электронной культуре [Текст] / А.Ю. Алексеев // Вузы культуры и искусств в мировом образовательном пространстве: новые пути наук о культуре: Материалы международной научно-практической конференции. Ч.2 (Москва, 28-29 мая 2009 года): сборник статей/ Научно-исследовательский институт МГУКИ. – М.: МГУКИ, 2009. – 246 с., С. 6-16
- [Алексеев, Бахтизин, 2009-8] Алексеев А.Ю., Бахтизин А.Р. Проблемы проекта искусственной личности //Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 03 июня 2009 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2009. URL : <http://www.scmal.ru/seminar/1-38>
- [Алексеев, 2010-1] Алексеев, А.Ю. Четыре лика электронной культуры [Текст] / А.Ю. Алексеев // Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2010. – 180 с., С. 50-68
- [Алексеев, 2010-2] Алексеев, А.Ю. Проблема сознания в электронной культуре [Текст] / А.Ю. Алексеев // Полигнозис, 3(39), 2010, С.129-141
- [Алексеев 2010-3] Алексеев. А.Ю. Роль нейрокompьютера в электронной культуре [Текст] / А.Ю. Алексеев // Журнал: Нейрокompьютеры: разработка и применение, № 8, 2010, С.14-26
- [Алексеев, 2010-4] Алексеев, А.Ю. (ред.) Электронная культура: феномен неопросветительства. [Текст] / А.Ю. Алексеева, С.Ю. Карпук – [ред.] – М.: Изд-во МГУКИ, 2010. – 180 с.
- [Алексеев, 2010-5] Алексеев, А.Ю. Когнитивные исследования в электронной культуре [Текст] / А.Ю. Алексеев // Материалы Второй международной научно-практической конференции "Актуальные проблемы современной когнитивной науки" – Иваново: Изд-во Ивановского госуниверситета, 2010. – 240 с. – С.102-108
- [Алексеев, 2011-1] Алексеев, А.Ю. Корсаков С.Н. – основатель когнитивной науки [Текст] / А.Ю. Алексеев, Фатеева Л.А. // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. – 350 с. – С. 124-131
- [Алексеев, 2011-2] Алексеев, А.Ю. Развитие проблематики сознания в электронной культуре [Текст] / А.Ю. Алексеев // Науки о культуре: инновации и приоритетные разработки: [сборник статей преподавателей МГУКИ] : в 2 ч./науч.ред.

- Л.С. Жаркова, В.М. Беспалов, В.Г. Иванова. – М.: МГУКИ, 2011. – Ч.1. – 176 с. – С.134-146
- [Алексеев, 2011-3] Алексеев, А.Ю. Роль комплексного теста Тьюринга в методологии искусственных обществ [Текст] / А.Ю. Алексеев // Ежеквартальный Интернет – журнал «Искусственные общества», Том 6, номер 1-4, I-IV квартал 2011 г. - С. 18-64
- [Алексеев, 2011-4] Алексеев, А.Ю. Проблемы электронной культуры: когнитивные исследования в образовательной среде [Текст] / А.Ю. Алексеев // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. – 350 с. – С. 160-170
- [Алексеев, 2011-5] Алексеев, А.Ю. Идея Комплексного Теста Тьюринга [Текст] / А.Ю. Алексеев // «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). – С. 8 - 19
- [Алексеев, 2011-6] Алексеев, А.Ю. Социокультурные параметры нейрокомпьютинга [Текст] / А.Ю. Алексеев // Полигнозис, 2(42), 2011, С.144-156
- [Алексеев, 2011-7] Алексеев, А.Ю. Нейрокомпьютер и творчество [Текст] / А.Ю. Алексеев, С.А. Игнатов, Т.А. Конькова // Журнал: Нейрокомпьютеры: разработка и применение, № 12, 2011, С. 3-18
- [Алексеев, 2011-8] Алексеев, А.Ю. Интеллектуализация электронной культуры – ключевое направление модернизации России [Текст] / А.Ю. Алексеев // Регионы России: Стратегии и механизмы модернизации, инновационного и технологического развития. Тр. Седьмой междунар. научн.-практ. конф. / РАН. ИНИОН. Отд. науч. сотрудничества и междунар. связей; Отв. ред. Ю.С. Пивоваров. – М., 2011. – Ч. 2. – 588 с. – С. 477 – 483
- [Алексеева, И.Ю. 1992] Алексеева И.Ю. Человеческое знание и его компьютерный образ. - М.: Наука, 1992. - 209 с.
- [Антопольский, 2010] Антопольский А.Б. Принципы создания единого информационного пространства вуза//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 82-86
- [Баксанский, 2011] Баксанский О.Е. Когнитивная наука и проблемное поле философии образования // Актуальные проблемы современной когнитивной науки. Материалы четвёртой всероссийской научно-практической конференции (с международным участием (20-21 октября 2011 года). Иваново: ОАО "Изд-во "Иваново", 2011. - 350 с. - С. 12-25
- [Баксанский, Кучер, 2005] Баксанский О.Е., Кучер Е.Н. Моделирование социального интеллекта в когнитивных науках //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелл, 2005. – 280 с. – С.120-124
- [Барендрегт, 1985] Барендрегт, Х. Ламбда-исчисление. Его синтаксис и семантика: Пер. с англ. — М.: Мир, 1985. — 606 с.

- [Бахтизин, 2005] Бахтизин А.Р. Моделирование социально-экономической системы с помощью технологии искусственного интеллекта //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.232-237
- [Бахтизин, 2006] Бахтизин А.Р. Искусственный интеллект и его приложения в экономико-математических моделях //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.347-378
- [Бахтизин, 2009] Бахтизин А.Р. Подход к разработке агент–ориентированных моделей// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Бахтизин, 2011] Бахтизин А.Р. Агентное моделирование на базе технологий искусственного интеллекта// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 281-310
- [Бескова, 2005] Бескова И.А. Вариант прочтения интенциональной семантики //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.190-192
- [Бирюков, 1989] Бирюков Б.В. Кибернетика, информатика, вычислительная техника, автоматика: проблемы становления и развития. Вклад естественной науки// Кибернетика: прошлое для будущего. Этюды по истории отечественной кибернетики. Теория управления. Автоматика. Биокибернетика. М., 1989.
- [Бирюков, 2004] Бирюков Б.В. Жар холодных чисел и пафос бесстрастной логики. Формализация мышления от античных времен до эпохи кибернетики. Изд. 3-е, переработанное и дополненное. М., 2004.
- [Бодякин, 2005] Бодякин В.И. Концепция построения искусственного разума на базе нейросемантического подхода //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.192-195
- [Брингсйорд и др., 2006] С.Брингсйорд, П.Беллоу, Д.Феруччи. Творчество, тест Тьюринга и улучшенный тест Лайвлейс. Пер.с англ. А. Ласточкина в кн. //Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 129 с. – С. 62-83 – 120 с.
- [Брукшир, 2004] Брукшир, Дж. Информатика и вычислительная техника. 7-е изд. - СПб.: Питер, 2004. - 620 с: ил.
- [Будущее ИИ, 1991] Будущее искусственного интеллекта. - М.: Наука, 1991. – 302 с.
- [Бульчев, 2000-1] Бульчев И.И. Категория "субъект" в свете методологии универсального логико-философского алгоритма. Вестник Тамбовского университета. Серия: Гуманитарные науки. 2000. № 2. С. 3-18.
- [Бульчев, Шутов, 2002] Бульчев И.И., Шутов Р.В. О сущности естественного и искусственного интеллектов. Вестник Тамбовского университета. Серия: Гуманитарные науки. 2002. № 1. С. 21-26.

- [Булычёв, Кирюшин, 2007] Булычёв И.И., Кирюшин А.Н. О базисных логико-философских константах теории коммуникации. *Философия и общество*. 2011. № 4. С. 97-116.
- [Васюков, 2006] Васюков В.Л. Формальная онтология и искусственный интеллект // *Искусственный интеллект: междисциплинарный подход*. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.305-324
- [Вельтман, 2002] Вельтман Ким. Электронная культура: достижения и перспективы. Пер. с англ. Н.В. Браккер // *Информационное общество*, 2002, вып. 1, сс. 24-30.
- [Вильянуэва, 2006] Вильянуэва Э. Что такое психологические свойства? *Метафизика психологии*. М.: Идея-Пресс, 2006. – 256 с.
- [Винер, 1983] Винер, Н. Кибернетика, или управление и связь в животном и машине. 2-е издание. М.: Наука; Главная редакция изданий для зарубежных стран, 1983. – 344 с.
- [Глазунов, 2010] Глазунов В.А. Методологические проблемы робототехники. Междисциплинарность, бифуркации, многокритериальность. Антропоморфные и параллельные роботы // *Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара*, 20 янв. 2010 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2010 г.. URL : <http://www.scm.ai.ru/seminar/1-41>
- [Глазунов, Алексеев 2007] Глазунов В.А., Алексеев А.Ю. Методологические вопросы робототехники // *Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара*, 25 апр. 2007 г., М.:ЦЭМИ РАН / НСМИИ РАН, 2007 г.. URL : <http://www.scm.ai.ru/seminar/1-19>
- [Глазунов, Глазунова, 2005] Глазунов В.А., Глазунова О.В. Моделирование творческого акта средствами робототехники // *Новое в искусственном интеллекте. Методологические и теоретические вопросы*. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.91-92
- [Грэхэм, 1991] Грэхэм Л.Р. Естествознание, философия и науки о человеческом поведении в Советском Союзе: Пер. с англ. – М. Политиздат, 1991. – 480 с.
- [Гук, 2010] Гук А.А. Утилитарно-практическое и художественно-эстетическое в экранном творчестве // *Электронная культура: феномен неопросветительства*. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 133-139
- [Декарт, 1989] Декарт Р. Сочинения в 2 т.: Пер. с лат. и франц. Т. I / Сост., ред., вступ. ст. В. В. Соколова. - М.: Мысль, 1989. – 654 с. – С. 315 (см. «Первоначала философии» - С.297- 422, пер. с лат. С. Я. Шейнман-Топштейн, с франц. Н. Н. Сретенского)
- [Делицын, 2010] Делицын Л.Л. Элементы префигуративного типа культуры в феномене неопросветительства: способ преодоления внутреннего цифрового разрыва // *Электронная культура: феномен неопросветительства*. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 158-163
- [Деннетт, 2004] Деннетт Дэниел С. Виды психики: на пути к пониманию сознания. Перевод с англ. А. Веретенникова. Под общ.ред. Л.Б. Макеевой. – М.: Идея-Пресс, 2004. – 184 с. – С.

- [Дмитриевская, 1995] Дмитриевская И.В. Принцип интенциональности в логике Г.Фреге // *Философия языка и семиотика*. Иваново, 1995
- [Дубровский, 1971] Дубровский Д.И. Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М., «Наука», 1971
- [Дубровский, 1980] Дубровский Д.И. Информация, сознание, мозг. М., «Высшая школа», 1980
- [Дубровский, 1994-1] Дубровский Д.И. Психика и мозг. Результаты и перспективы исследований. // *Мозг и разум*. М.: Наука, 1994 – 176 с.
- [Дубровский, 1994-2] Дубровский Д. И. Обман. Философско-психологический анализ. М., 1994
- [Дубровский, 2002-1] Дубровский Д.И. Проблема идеального. Субъективная реальность. — М., 2002. — 368 с.
- [Дубровский, 2002-2] Дубровский Д.И. Новая реальность: человек и компьютер / *Полигнозис* № 3 (23), С. 20-32
- [Дубровский, 2003-1] Дубровский Д.И. Проблема духа и тела: возможности решения (в связи со статьей Т.Нагеля «Мыслимость невозможного и проблема духа и тела»). *Вопросы философии*, 2003 г.
- [Дубровский, 2003-2] Дубровский Д.И. Новое открытие сознания? (По поводу книги Дж. Серла «Открывая сознание заново») // *Вопросы философии*, 2003, №7
- [Дубровский, 2004-1] Дубровский Д.И. Тест Тьюринга, парадигма функционализма и проблема сознания// *Методологические и теоретические аспекты искусственного интеллекта*. Материалы студенческой конференции “Философия искусственного интеллекта”, МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева – М.: МИЭМ-ИИнтелЛ, 2006. – 224 с. – С. 126-128 – С. 109
- [Дубровский, 2004-2] Дубровский Д.И. Гносеология субъективной реальности. К постановке проблемы // *Эпистемология и философия науки*, 2004, № 2.
- [Дубровский, 2005-1] Дубровский Д.И. 2005. Искусственный интеллект и проблема сознания // *Философия искусственного интеллекта*. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.26-32
- [Дубровский, 2005-2] Дубровский Д.И. Сознание, мозг, искусственный интеллект//*Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара*, 18 мар 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmal.ru/seminar/1-1>
- [Дубровский, 2005-3] Дубровский Д.И. 2005. Искусственный интеллект и проблема сознания // *Новое в искусственном интеллекте. Методологические и теоретические вопросы*. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.42-48
- [Дубровский, 2006-1] Дубровский Д.И. Бессознательное (в его отношении к сознательному) и квантовая механика. *Философские науки*, № 8, 2006. – С. 52 - 73
- [Дубровский, 2006-2] Дубровский Д.И. 2006. Приветственное слово участникам конференции // *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. – С.8-11
- [Дубровский, 2006-3] Дубровский Давид 2006. Тест Тьюринга, парадигма функционализма и проблема сознания. Подведение итогов работы секции № 1 //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. – С.93-97
- [Дубровский, 2006-4] Дубровский Д.И. Сознание, мозг, искусственный интеллект //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.75-89
- [Дубровский, 2007] Дубровский Д. И. Сознание, мозг, искусственный интеллект: сб. статей. – М.: 2007. – 272 с.
- [Дубровский, 2009-1] Дубровский Д.И. (ред.) Проблемы сознания в философии и науке. Под ред. проф. Д.И. Дубровского. – М.: «Канон+» РООИ «Реабилитация», 2009. – 472 с., С.195-214
- [Дубровский, 2009-2] Дубровский Д.И. Проблема сознания: опыт обзора основных вопросов и теоретических трудностей //«Проблемы сознания в философии и науке» / под ред. проф. Д.И. Дубровского. – М.: «Канон+»РООИ «Реабилитация», 2009. – 472 с., С.8-44
- [Дубровский, 2009-3] Дубровский Д.И. Мозг: фундаментальные и прикладные проблемы. Философские подходы к проблеме «мозг и психика»: к вопросу о расшифровке нейродинамических кодов явлений субъективной реальности // Доклад на научной сессии Общего собрания Российской Академии Наук 15 декабря 2009 г., Президиум РАН, г.Москва
- [Дубровский, 2010] Дубровский Д.И. Субъективная реальность, мозг и развитие НБИК-конвергенции: эпистемологические проблемы// Эпистемология вчера и сегодня/ Рос.акад.наук, Ин-т философии; Отв.ред. В.А. Лекторский. - М.: ИФРАН, 2010. – 188 с. – С. 78
- [Дубровский, 2011-1] Дубровский Д.И. Актуальные проблемы интерсубъективности// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 129-148
- [Дубровский, 2011-2] Дубровский Д.И. Трудная проблема сознания (в связи с книгой В.В.Васильева)// Вопросы философии, № 9, 2011, С. 136-148. URL : <http://elibrary.ru/item.asp?id=16860687>
- [Дубровский, Лекторский, 2006] Дубровский Д.И., Лекторский В.А. 2006. От редакторов //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.3
- [Дэвидсон, 1990] Дэвидсон Д. Что означают метафоры // Теория метафоры. М., 1990. С. 172-193. Пер. М.А. Дмитриевской. URL : <http://kant.narod.ru/davidson.htm>
- [Егоров, 2005] Егоров А.Г. Информационно-поисковая система экспертного типа по гуманитарным наукам как аналог машины Раймонда Луллия //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.339-343

- [Жижек, 2002] Жижек С., 2002. Обойдемся без секса, ведь мы же пост-люди! Пер. с англ. Артема Смирнова. URL : http://www.ruthenia.ru/logos/kofr/2002/2002_12.htm#_edn1#_edn1
- [Зотов, 2005] Зотов А.Ф. 2005. Что значит «мыслить»? //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.163-167
- [Зотов, 2006] Зотов А.Ф. Роль феноменологии в организации междисциплинарных исследований в области искусственного интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.22-32
- [Иваницкий, 2006-1] Иваницкий А.М. Думающий мозг и искусственный интеллект – встречное движение // Труды XXV межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем. Ч. 4. – Серпухов: Серпуховский ВИ РВ, 2006 г. – 303 с.
- [Иваницкий, 2006-2] Иваницкий А.М. Физиологические основы сознания и проблема искусственного интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.90-99
- [Иванов, 1999] Иванов Е.М. К проблеме «вычислимости» функции сознания, 1999; <http://filosof.historic.ru/books/item/f00/s00/z0000899/index.shtml>
- [ИИС, 1991] Интеллектуальные информационные системы. М.: Военное издательство, 1991
- [ИИ, межд.подход, 2006] Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с.
- [ИИ: философия, методология, инновации, 2006] 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.: ИИнтелЛЛ, 2006. – 440 с.
- [ИИ, 1990] Искусственный интеллект – В 3-х кн. Кн.2. Модели и методы: Справочник/Под ред. Д.А. Поспелова – М.: Радио и связь, 1990. – 304 с.
- [Истратов, 2009] Истратов В.А. «Агентно-ориентированная модель поведения человека в социально-экономической среде». Автореферат диссертации, представленной на соискание ученой степени кандидата экономических наук по специальности 08.00.13. – Математические и инструментальные методы экономики. М: ЦЭМИ РАН, 2009
- [Калинин, 2011] Калинин П.Е. Становление статико-динамичного и континуально-дискретного единства в деятельности сознания (онтологический и гносеологический аспекты). Автореферат на соискание ученой степени кандидата философских наук по специальности 09.00.01 – онтология и теория познания. – Иваново: ИГУ, 2011.
- [Колин, 2005] Колин К.К. Философские основы информатики //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.248-252.

- [Колин, 2006-1] Колин К.К. Россия и мир на пути к информационному обществу//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 01 фев 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scmair.ru/seminar/1-8>
- [Колин, 2006-2] Колин К.К. Становление информационного общества в России и национальная безопасность //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.379-392
- [Колин, 2010-1] Колин К.К. Системная модернизация России и проблемы развития информационного общества//Государственная служба, № 2, 2010, С.32-37
- [Колин, 2010-2] Колин К.К. Философские проблемы информатики. – М.: БИНОМ. Лаборатория знаний. 2010. – 264 с.
- [Колин, 2011] Колин К.К. Электронная культура в информационном обществе// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 224-252
- [Колмогоров, 1964] Колмогоров А. Автоматы и жизнь // Возможное и невозможное в кибернетике. М., 1964
- [Конькова, 2009] Конькова Т.А. Электронная культура? Критика тотальной вычислимости// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Корсаков (Михайлов), 2009] Корсаков С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А.С. Михайлова. – М.: МИФИ, 2009. – 44 с.
- [Корсаков (Сыромятин), 2009] Корсаков С. Н. (1832 г.). Очерк о новом способе исследования посредством машин для сравнения идей. (Перевод с франц. А.В. Сыромятина) // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Кочергин, 2005] Кочергин А.Н. Искусственный интеллект и мышление //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.37-39
- [Кочергин, 2006] Кочергин А.Н. 2006. Искусственный интеллект, психика, творчество //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.209-222
- [Кубрякова, 1996] Краткий словарь когнитивных терминов. Под общ.редакцией Е.С. Кубряковой. М.: МГУ, 1996 – 245 с.
- [Кувакин, 1998] Твой рай и ад. Человечность и бесчеловечность человека: Философия, психология и стиль мышления гуманизма. —СПб.: «Алетейя», —М.: «Логос», 1998. — 360 с.
- [Кудряшова, 2005] Кудряшова Т.Б. Онтология языков познания. Ч.I. - Иваново: ОАО "Изд-во "Иваново", 2005. - 316 с.
- [Кудряшова, 2010] Кудряшова Т.Б. Пути возрождения инженерной культуры через развитие творческого когнитивного потенциала субъектов высшего образо-

- вания //Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 86-91
- [Кудряшова, 2011] Кудряшова Т.Б. Роль пространственного воображения в формировании и развитии способности к категоризации// Актуальные проблемы современной когнитивной науки. Материалы четвёртой всероссийской научно-практической конференции (с международным участием (20-21 октября 2011 года). Иваново: ОАО "Изд-во "Иваново", 2011. - 350 с. - С. 32-36
- [Кузичев, 2006] Кузичев А.С. Программа Колмогорова, интеллектуальные системы и теоремы Гёделя о неполноте //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.330-346
- [Кузнецов, 1989] Кузнецов В.Е. Представление в ЭВМ неформальных процедур: продукционные системы. М.: Наука. – 1989 – 160 с.
- [Кузнецов, 1991] Кузнецов В.Г. Герменевтика и гуманитарное познание. М.: Изд-во Моск.ун-та, 1991-192 с.
- [Кузнецов, Сергиенко, Холодная, 2006] Кузнецов О.П. Может ли искусственный интеллект развиваться независимо от исследований естественного интеллекта?//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 01 ноя 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scm.ai.ru/seminar/1-14>
- [Лахути, 2009] Лахути Д.Г. О пяти кругах искусственного интеллекта и о дискуссии Поппера с Бернайсом // Вопросы философии, № 10, 2009, С.116-120. URL : <http://elibrary.ru/item.asp?id=13214944>
- [Лейбниц, 1982] Лейбниц. В.Г. Сочинения в четырех томах: Том I. / Ред. и сост., авт. вступит. статьи и примеч. В.В.Соколов; перевод Я.М.Боровского и др. – М.: «Мысль», 1982. – 636 с.
- [Лекторский, 2005] Лекторский В.А. 2005. Философия, когнитивные науки и искусственный интеллект //Философия и будущее цивилизации: Тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т.1. - М.: Современные тетради, 2005. - 768 с. – С.728
- [Лекторский, 2006] Лекторский В.А. Философия, искусственный интеллект и когнитивная наука //Искусственный интеллект: междисциплинарный подход. Под ред.Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛЛ, 2006.– 448 с.– С.12-22
- [Лекторский, 2010] Лекторский В.А. Эпистемологическое исследование когнитивных процессов. Эпистемология вчера и сегодня/ Рос.акад.наук, Ин-т философии; Отв.ред. В.А. Лекторский. - М.: ИФРАН, 2010. – 188 с.
- [Лекторский, 2011] Лекторский В.А. Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 3-16
- [Лещёв, 2005] Лещёв С.В. Искусственное сознание и проблема интерпретации //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.143-145

- [Лещёв, 2010] Лещёв С.В. Археология электронной культуры: о харизме новых форм коммуникации//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С.54-58
- [Линдсей, Норман, 1974] П. Линдсей, Д. Норман. Переработка информации у человека. Введение в психологию. Пер. с англ. Под ред. А.Р. Лурия. М.: Мир, 1974 – 550 с.
- [Лобовиков, 2005] Лобовиков В.О. Естественное право как математика свободы и морально-правовое программирование автономных адаптивных роботов //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.105-107
- [Макаров, 2006] 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.5-12.
- [Макаров, 2010] Макаров В.Л. Социальный кластеризм. Российский вызов / М.: Бизнес Атлас, 2010. – 272 с.
- [Макаров, 2011] Макаров В.Л. Становится ли человеческое общество стабильнее?// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 211-223
- [Макаров, Бахтизин, 2005] Макаров В.Л., Бахтизин А.Р. Компьютерное моделирование искусственных миров//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 20 апр 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmi.ru/seminar/1-2>
- [Макеева, 1996] Макеева Л.Б. Философия Х.Патнэма. – М., 1996. – 190 с.
- [Марков, 1984] Марков А.А., Нагорный Н.М. Теория алгоритмов - М.: Наука, 1984
- [Мельчук, 1995] Мельчук И.А. Русский язык в модели «Смысл-Текст» - Москва-Вена: Школа «Языки русской культуры», 1995. – 682 с.
- [Меркулов, 2005] Меркулов И.П. Современная эпистемология: синтез информационных и эволюционных представлений //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.211-215
- [Меркулов, 2006] Меркулов И.П. Эпистемологический анализ когнитивной природы математических и логических формализмов //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.289-304
- [Микешина, Опенков, 1997] Микешина Л.А., Опенков М.Ю. Новые образы познания и реальности. – М.: Российская политическая энциклопедия (РОССПЭН), 1997. – 238 с.
- [Миронов, 2006] Миронов В.В. Информационное пространство: диалог культур //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.392-398
- [Миронов, 2011] Миронов В.В. Размышления о реформе российского образования: Доклад на международной научной конференции «Философия и образование в процессе трансформации культуры», посвящённой 70-летию воссоздания

- философского факультета в структуре МГУ имени М.В. Ломоносова. – М.: Издатель Воробьев А.В., 2011. – 64 с.
- [Михеенкова, Финн, 2005] Михеенкова М.А., Финн В.К. Проблемы создания интеллектуальных систем для анализа социального поведения //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.366-368
- [Михеенкова, Финн, 2010] Михеенкова В.А., Финн В.К. Интеллектуальный анализ данных и проблемы когнитивной социологии//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 24 фев 2010 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2010 г.. URL : <http://www.scmair.ru/seminar/1-42>
- [Моисеев, 2005] Моисеев В.И. Интервал Тьюринга и имитация интеллекта //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2005. – 280 с. – С.215-217
- [Нагуманова, 2007] Нагуманова С.Ф. Аргумент мыслимости против материализма// Философские науки, № 3, 2007 г.
- [Нариньяни, 2005-1] Нариньяни А.С. Путь России в информационной технологии XXI века. // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2005. – 280 с. – С. 11-25
- [Нариньяни, 2005-2] Нариньяни А.С. «НОМО – два в одном (Homo Sapience в ближайшей перспективе) //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2005. – 280 с. – С.259-278
- [Нариньяни, 2005-3] Нариньяни А.С. Перспективы развития интеллектуальных информационных технологий нового поколения//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 май 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmair.ru/seminar/1-3>
- [Нариньяни, 2006] Нариньяни А.С. Между Знанием и Незнанием – Наивная Топография 2 //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2006. – 448 с. – С.49-74
- [Новое в ИИ, 2005] Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2005. – 280 с.
- [Огурцов, 2005] Огурцов А.П. 2005. Достижения и трудности в моделировании интеллектуальных актов //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.56-60
- [Огурцов, 2006] Огурцов А.П. Возможности и трудности в моделировании интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛЛ, 2006. – 448 с. – С.32-48
- [Осипов, 2010] Осипов Г.С. Основные направления и тенденции развития работ в области искусственного интеллекта в Европе//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного науч-

- но-теоретического семинара, 29 окт 2010 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2010 г.. URL : <http://www.scmair.ru/seminar/1-44>
- [Пассмор, 2002] Пассмор, Дж. Современные философы. Перевод с англ. Л.Б.Макеевой. - М.: Идея-Пресс, 2002. - 192 с.
- [Пенроуз, 2005] Пенроуз Р. Новый ум короля: о компьютерах, мышлении и законах физики: Пер. с англ./ Общ.ред. В.О. Малышенко – М.: Едиториал УРСС, 2005 – 400 с.
- [Петренко, Осипов, 2005] Петренко В.Ф., Осипов Г.С. Междисциплинарный подход к исследованию сознания и личности: экспериментальная психосемантика//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 05 окт 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmair.ru/seminar/1-5>
- [Петрунин, 1996] Петрунин Ю.Ю. От тайного знания к нейрокомпьютеру: Очерки по истории искусственного интеллекта. М.: Университетский гуманитарный лицей. 1996
- [Петрунин, 2001] Петрунин Ю.Ю. Искусственный интеллект//Новая философская энциклопедия в 4 томах. М. «Мысль». 2001. Т. 2.
- [Петрунин, 2002] Петрунин Ю.Ю. Искусственный интеллект: история, методология, философия. М.: Звездопад. 2002
- [Петрунин, 2010] Петрунин Ю.Ю. Модели и методы искусственного интеллекта в управлении знаниями// Государственное управление. Электронный вестник. Выпуск № 24. Сентябрь 2010 (в соавторстве). URL : http://e-journal.spa.msu.ru/24_2010Petrinin_Petrulina.html
- [Петрунин, 2010] Петрунин Ю.Ю., Рязанов М.А., Савельев А.В. Философия искусственного интеллекта в концепциях нейронаук. М.: МАКС Пресс, 2010.–83с.
- [Петрунин, Алексеев, 2011] Петрунин Ю.Ю., Алексеев А.Ю. Искусственный интеллект как феномен современной культуры//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 08 дек 2011 г., г.Москва: ИФ РАН / НСМИИ РАН, 2011 г.. URL : <http://www.scmair.ru/seminar/1-50>
- [Поваров, 2005-1] Поваров Г.Н. С.Н. Корсаков – русский пионер искусственного разума //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.198-199
- [Поваров, 2005-2] Поваров Г.Н. Диалектический материализм и проблема искусственного интеллекта //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.196-198
- [Полосухин, 1993] Полосухин Б.М. Феномен вечного бытия. Некоторые итоги размышлений по поводу алгоритмической модели сознания. – М.:Наука, 1993 – 176 с.
- [Портнов, 1994] Портнов, А.Н. Язык и сознание: основные парадигмы исследования проблемы в философии XIX-XX вв. Издательство: Иваново: Ивановский Государственный Университет, 1994. - 336 с.
- [Поспелов, 1982] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту – М.: Наука, 1982. – 224 с., С. 5- 42

- [Рассел, 1999] Рассел Бертран. Исследование истины и значения. Пер. Ледникова Е.Е., Икифорова А.Л. – М.: Идея-Пресс, Дом интеллектуальной книги, 1999. – 440с.
- [Рассел, Норвиг 2006] Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд. – М.: Вильямс, 2006. – с. 1408.
- [Реале, Антисери, 1997] Реале Дж., Антисери Д. Западная философия от истоков до наших дней. Том 2. Средневековье. – ТОО ТК «Петрополис», 1997. – 368 с. – С.206 – 207.
- [Редько, 2006] Редько В.Г. Задача моделирования когнитивной эволюции //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.243-258
- [Редько, 2009] Редько В.Г. Об исследовании когнитивной эволюции// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С. 99-104
- [Резник, 1999-1] Резник Ю. М. Введение в социальную теорию. Социальная онтология. Пособие. — М.: Институт востоковедения РАН, 1999. — 514с.
- [Резник, 1999-2] Резник Ю. М. Введение в социальную теорию. Социальная эпистемология. Пособие. — М.: Институт востоковедения РАН, 1999. — 327 с.
- [Резник, 2003] Резник Ю. М. Введение в социальную теорию: Социальная системология / Ю. М. Резник; Ин-т человека. — М.: Наука, 2003. — 525с.
- [Розин, 2001] Розин Вадим. Семиотические исследования. – М.: ПЕР СЭ; СПб: Университетская кни-га, 2001 - 256 с.
- [Розин, 2005] Розин В.М. 2005. Мышление и искусственный интеллект //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.157-164
- [Розин, 2011] Розин В.М. Концепция искусственного интеллекта и общество как эвристический источник идей для нейрокомпьютера нового поколения// Нейрокомпьютер, № 1, 2011 г., С.4-10
- [Рыжов, 2005] Рыжов В.П. Художественное творчество как информационный процесс //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.254-267
- [Савельев, 2004] Савельев А.В. Нейрокомпьютеры в изобретениях // Нейрокомпьютеры: разработка и применение, 2004, № 2-3, С. 33-49
- [Савельев, 2005] Савельев А.В. Искусственный интеллект или искусственный социум? //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.210-212
- [Савельев, 2006] Савельев А. В. Самоорганизационное нейроуправление на реверберационных нейропроцессах // в сб. тр. МИФИ: "Нейроинформатика-2006", М.: МИФИ, 2006, ч. I, с. 172-180
- [Савельев, 2008] Савельев А.В. Нейрологические аспекты клеточной нейроматематики // журнал “Искусственный интеллект”, НАН Украины, Донецк, 2008, № 4, с. 612-623

- [Савельев, 2009] Савельев А.В. Нейроэкзистенциальное моделирование распространения спайков – путь к новой концепции субстрата долговременной нейронной памяти // журнал «Искусственный интеллект» НАН Украины, г. Донецк, 2009 г. С.329-335
- [Савельев, 2010] Савельев А.В. Каузальность информационной культуры и «нейрокомпьютеризация» сознания // Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С.46-53
- [Самохвалова, 2005] Самохвалова В.И. Предполагает ли машинное творчество возможность машинного вдохновения? // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.62-66.
- [Самохвалова, 2006] Самохвалова В.И. «Человеческое, слишком человеческое», или процессор в экстазе // Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.423-442.
- [Самохвалова, 2007] Самохвалова В.И. Творчество: божественный дар; космический принцип; родовая идентичность человека: Научное издание. – М.: РУДН, 2007. – 538 с.
- [Сергиенко, Холодная, 2006] Сергиенко Е.А., Холодная М.А. Истоки познания: революция в когнитивной психологии развития // Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 29 мар 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scm.ai.ru/seminar/1-10>
- [Серёдкина, 2007] Серёдкина Е.В. Общая теория рациональности и артефакты в проекте OSCAR Дж. Поллока. // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн. ун-та, 2007. – 210 с. – С. 108-122.
- [Серль, 2002] Серль Дж. Открывая сознание заново. Пер. с англ. А.Ф. Грязнова. М.: 2002. – 256 с.
- [Серль, 2006] Серль Джон Р. Разумы, мозги, программы (Перевод с англ. Д. Родионова) // Тест Тьюринга. Роботы. Зомби. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.6–20.
- [Скворцов, 2005] Скворцов А.А. Нравственные ценности виртуального мира // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.164-166.
- [Скворцов, 2006] Скворцов А.А. 2006. Мировоззренческие и теоретические основы этических исследований виртуальности // Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.411-433.
- [Сломэн, 2006] Сломэн Арон. 2006. Что значит быть камнем? (Перевод В. Крючкова) // Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.86-102.
- [Смирнов Г.С., 1998] Смирнов Г.С. Ноосферное сознание и ноосферная реальность: философские проблемы ноосферного универсума. Иваново, 1998

- [Смирнов Д.Г., 2008] Смирнов Д.Г. Семиософия ноосферного универсума [Текст] : ноосфера и семиосфера в глобальном дискурсе / Д. Г. Смирнов ; [науч. ред. А. Н. Портнов] ; М-во образования и науки Рос. Федерации, Федер. агентство по образованию, ГОУ ВПО "Иван. гос. ун-т". - Иваново : Иван. гос. ун-т, 2008. - 370 с.
- [Смирнова, 2005-1] Смирнова Н.М. Трансцендентально-феноменологическая теория конституирования смыслов и ее эвристические возможности для философии искусственного интеллекта // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.217-220.
- [Смирнова, 2005-2] Смирнова Н.М. Методология феноменологического конституирования и её значение для философии искусственного интеллекта // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛ, 2005. – 280 с. – С.166-169
- [Смирнова, 2006] Смирнова Н.М. Трансцендентальная intersубъективность, проблема «чужих сознаний», искусственный интеллект // В кн. Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИИнтелЛ, 2006. – 448 с. – С. 163 – 179.
- [Смирнова, 2011] Смирнова Н.М. Intersубъективность и проблема «Других сознаний» // Естественный и искусственный интеллект: методологические и социальные проблемы / Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 106-128.
- [Стёпин, 1991] Стёпин В.С. Научная рациональность в гуманистическом измерении. // В кн. О человеческом в человеке / Под общ.ред. И.Т.Фролова – М., 1991. – 384 с.
- [Стёпин, 1992] Стёпин В.С. Философская антропология и философия науки. - М.: Высшая школа, 1992. - 191 с.
- [Стёпин, 2007] Стёпин В.С. Постнеклассическая рациональность и информационное общество // Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 31 окт 2007 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2007 г.. URL : <http://www.scm.ai.ru/seminar/1-22>
- [Тест Тьюринга, 2006] Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с.
- [Трушкина, 2005] Трушкина Н.Ю., Михейкин И.В., Кундин Л.В. Решение проблемы сознания в рамках концепции сильного искусственного интеллекта // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.79-82
- [Трушкина, 2006] Трушкина Н.Ю. «Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта». Диссертация на соискание ученой степени кандидата философских наук по специальности 09.00.08. – философия науки и техники. М.: МГУ им. М.В. Ломоносова
- [Тьюринг, 1960] Алан М. Тьюринг. Может ли машина мыслить? (Под ред. Б.В. Бирюкова). М., 1960. С. 19-158
- [Уёмов, 1963] Уёмов А.И. Вещи, свойства и отношения. М., 1963

- [Уитби, 2004] Уитби Б. Искусственный интеллект: реальна ли Матрица. – М.: ФАИР-ПРЕСС, 2004. – с. 224.
- [Управление информация интеллект, 1976] Управление, информация, интеллект. Под ред. А.И.Берга и др. М.: Мысль, 1976.–383 с.– С. 282.
- [Философия и будущее цивилизации, 2005] Алексеев, А.Ю., Глазунов, В.А., Лекторский В.А. (ред.). Круглый стол «Философско-методологические проблемы когнитивных и компьютерных наук»// Философия и будущее цивилизации: Тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т.1. – М.: Современные тетради, 2005. – 768 с.
- [Философия ИИ, 2004] 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.
- [Философия ИИ, 2005] Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с.
- [Философско-методологические проблемы ИИ, 2007] Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара/ Под ред. Е.В.Серёдкиной. - Пермь: Изд-во Перм.гос.техн.ун-та, 2007. - 210 с.
- [Финн, 2007] Финн В.К. Интеллектуальные системы и общество: Сборник статей/ Предисл. Д.А. Поспелова, Д.Г. Лахути, В.Б. Тарасова. – М.: КомКнига, 2007. – 352 с. – С.30
- [Финн, 2009] Финн В.К. К структурной когнитологии: феноменология сознания с точки зрения искусственного интеллекта //Вопросы философии №1, 2009
- [Фланеган, Полджер, 2006] Фланеган Оуэн, Полджер Томас. 2006. Зомби и роль сознания (Перевод Т. Кураевой) //Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.103-116
- [Холодная, Сергиенко, 2005] Холодная М.А., Сергиенко Е.А. Структура и функции естественного интеллекта в контексте проблемы искусственного интеллекта//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 ноя 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scm.ai.ru/seminar/1-6>
- [Хофштадтер, 2001] Хофштадтер Д.Р. Гёдель, Эшер, Бах: эта бесконечная гирлянда. – Самара: Издательский Дом «Бахрах-М», 2001 – 752 с.
- [Цанн-кай-си, 2005] Цанн-кай-си Ф.М., Плеханов Е. Гуманизм как философско-антропологическая проблема//Здравый смысл, лето 2005 № 3 (36)
- [Чарнецки, Айзенкер, 2005] Чарнецки К., Айзенкер У. Порождающее программирование: методы, инструменты, применения. Для профессионалов. - СПб.: Питер, 2005. - 731 с.
- [Черниговская, 2006] Черниговская Т.В. 2006. Зеркальный мозг, концепты и язык: цена антропогенеза //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛ, 2006. – 448 с. – С.127-148
- [Шалютин, 1985] Шалютин, С.М. Искусственный интеллект [Текст] : гносеологический аспект / С. М. Шалютин. - Москва : Мысль, 1985. - 198

- [Шалютин, 2006] Шалютин С.М. Машины, Люди. Ценности.- Курган: Изд-во Курганского гос.университета, 2006 -393 с.
- [Шлыкова, 2009] Шлыкова О.В. Электронная культура: дефиниции и тенденции развития//Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. - С.30-36
- [Шульга, 2002] Шульга Е.Н. Когнитивная герменевтика, М., ИФ РАН, 2002
- [Шульга, 2008] Шульга Е.Н. Понимание и интерпретация. М.: Наука, 2008
- [Шэнк, 1980] Шэнк Р. Обработка концептуальной информации. М.: Энергия, 1980
- [Электронная культура, 2009] Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Юлина, 2004] Юлина Н.С. Головоломки проблемы сознания: концепция Дэниела Деннета. – М.: Канон+, 2004. – 544 с. – С. 168
- [Юлина, 2005] Юлина Н.С. 2005. Обретение разумности: «сократический диалог» и интеракция с компьютером //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.82-85
- [Юлина, 2006] Юлина Н.С. 2006. Улучшение качества мышления: «сократический диалог» и интеракция с компьютером //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.180-193
- [Янковская, 2008] Янковская Е.А. Гетерархический принцип устройства познавательного опыта. Автореферат на соискание ученой степени кандидата философских наук. – Иваново, ИГУ, 2008
- [Янковская, 2011] Янковская Е.А. Гетерархическая модель интересубъективного взаимодействия // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. – 350 с.
- [Ярославцева, 2009] Ярославцева, Е.И. Искусственный интеллект в цифровом мире// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С. 48-54
- [Ясницкий, 2005] Ясницкий Л.Н. Введение в искусственный интеллект. - М.: Издательский центр «Академия», 2005. – 176 с.
- [Abelson, 1968] Abelson, RP. 1968, Simulation of Social Behavior, in G. Lindzey and E. Aronson, eds. Handbook of Social Psychology Reading, MA.: Addison Wesley, pp. 274–356.
- [Balog. 1999] Balog, K. 1999. Conceivability, possibility, and the mind-body problem. Philosophical Review 108:497-528.. URL : <http://humanities.ucsc.edu/NEH/balog.htm>
- [Barresi, 1987] Barresi, J. 1987. Prospects for the Cyberiad: Certain Limits on Human Self-Knowledge in the Cybernetic Age, Journal for the Theory of Social Behavior 17, pp. 19–46.
- [Barresi, 1990] Barresi, J. 1999. On becoming a person. Philosophical Psychology 12:79-98.

- [Barresi, 1995] Barresi, John (1995) Building Persons: Some Rules for the Game, *Psychology*: 6,#12 Robot Consciousness (7). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000459/>
- [Blackmore, 2003] Blackmore, Susan J. 2003. *Consciousness : An Introduction*. US Edition Published by Oxford University Press, New York 2004. URL : <http://www.susanblackmore.co.uk/Books/Consciousness/Consciousness%20Chap%202.pdf>
- [Block, 1978] Block, N. 1978. Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9:261-325. Reprinted in *Readings in the Philosophy of Psychology* (MIT Press, 1980).. URL : <http://perso.univ-rennes1.fr/pascal.ludwig/troubles.pdf>
- [Block, 1980] Block, N. 1980. Functionalism. In (N. Block, ed) *Readings in the Philosophy of Psychology*, Vol. 1. MIT Press. (a revised version of the entry on functionalism in *The Encyclopedia of Philosophy* Supplement, Macmillan, 1996). URL : <http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.pdf>
- [Block, 1981-1] Block, N. 1980. Troubles with functionalism. In (N. Block, ed) *Readings in the Philosophy of Psychology*, Vol 1. Harvard University Press.. URL : <http://perso.univ-rennes1.fr/pascal.ludwig/troubles.pdf>
- [Block, 1981-2] Block, N. 1981, Psychologism and Behaviorism, *Philosophical Review* 90, pp. 5-43.
- [Bringsjord, 1994] Bringsjord, Selmer (1994) What Robots Can and Can't be , *Psychology*: 5,#59 Robot Consciousness (1). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000418/#html>
- [Bringsjord, 1996] Bringsjord, S. 1996, The Inverted Turing Test is Provably Redundant. *Psychology* 7(29).. URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.29>.
- [Bringsjord, 1999] Bringsjord, S. 1999. The zombie attack on the computational conception of mind. *Philosophy and Phenomenological Research* 59:41-69.
- [Bringsjord, 2001-1] Bringsjord, S. 2001 The Zombie Attack on the Computational Conception of Mind. URL : <http://www.rpi.edu/~brings/SELPAP/zombies.ppr.pdf>
- [Bringsjord, 2001-2] Bringsjord, S. 2001. Creativity, the Turing test, and the (better) Lovelace test. *Minds & Machines* 11:3-27.. URL : <http://www.rpi.edu/~faheyj2/SB/SELPAP/DARTMOUTH/lt3.pdf>.
- [Butler, 1865] Butler, Samuel. Erehwon, London, 1865. (Chapter 23-25: The Book of the Machines). URL : <http://www.hoboes.com/FireBlade/Fiction/Butler/Erehwon/erewhon23/>
- [Chaitin, 2003] Chaitin, G.J. *Algorithmic Information Theory*, Cambridge University Press, 1987, Third Printing, 2003, 234 p.
- [Chalmers, 1993] Chalmers, D. J. 1993. The Conscious Mind: In Search of a Fundamental Theory. URL : <http://jamaica.u.arizona.edu/~chalmers/book/tcm.html>
- [Chalmers, 2011] Chalmers, D. J. *Zombies on the web*. URL : <http://jamaica.u.arizona.edu/~chalmers/zombies.html>
- [Chalmers, Bourget, 2010] Chalmers, D. Bourget. A bibliography of work in the philosophy of mind, the philosophy of cognitive science. URL : <http://consc.net/mindpapers>
- [Church, 1936] Church, Alonzo. An Unsolvability Problem of Elementary Number Theory *American Journal of Mathematics*, Vol. 58, No. 2. (Apr., 1936), pp. 345-363.. URL :

- <https://www.fdi.ucm.es/profesor/fraguas/CC/church-An%20Unsolvable%20Problem%20of%20Elementary%20Number%20Theory.pdf>
- [Churchland, 1996] Churchland, P. S. 1996. The hornswoggle problem. *Journal of Consciousness Studies* 3:402-8. Reprinted in Shear 1997,
- [Clifton, 2003] Clifton A., 2003. The Introspection game. Or, Does The Tin Man Have A Heart. URL : [@http://cogprints.ecs.soton.ac.uk/archive/00003483/01/Introspection_game](http://cogprints.ecs.soton.ac.uk/archive/00003483/01/Introspection_game)
- [Colby, 1981] Colby, K.M. 1981, Modeling a Paranoid Mind, *Behavioral and Brain Sciences* 4(4), pp. 515–560.
- [Colby, Hilf, Weber, 1971] Colby, K.M. Hilf, F.D. and Weber, S. 1971, Artificial Paranoia, *Artificial Intelligence* 2, pp. 1–25.
- [Colby, Hilf, Weber, Kraemer, 1972] Colby, K.M. Hilf, F.D., Weber, S. and Kraemer, 1972, Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes, *Artificial Intelligence* 3, pp. 199–222.
- [Cole, 2009] Cole, David. Chinese Room Argument. *Stanford Encyclopedia of Philosophy*, 2009. URL : <http://plato.stanford.edu/entries/chinese-room/>
- [Crick, 1994] Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. Scribners.
- [Cummins, 2000] Cummins, Robert. *Minds, brains, and computers : the foundations of cognitive science ; an anthology*. Malden, Mass. [u.a.] : Blackwell, 2000.. URL : <http://www.worldcat.org/title/minds-brains-and-computers-the-foundations-of-cognitive-science-an-anthology/oclc/246213097>
- [Dennett, 1984] Dennett, D. C. 1984. The Age of Intelligent Machines: Can Machines Think? In (M. Shafto, ed) *How We Know*. Harper & Row.. URL : <http://www.kurzweilai.net/articles/art0099.html?printable=1>
- [Dennett, 1985] Dennett, D. C. 1985. Can machines think? In *How We Know* (Shafto).
- [Dennett, 1994-1] Dennett, D. C. 1994. The practical requirements for making a conscious robot. *Philosophical Transactions of the Royal Society A* 349:133-46.
- [Dennett, 1995] Dennett, D. C. 1995. Cog: Steps toward consciousness in robots. In (T. Metzinger, ed) *Conscious Experience*. Ferdinand Schoningh.
- [Dennett 1995-1] Dennett Daniel C. The Unimagined Preposterousness of Zombies. Commentary on T. Moody, O. Flanagan and T. Polger, *Journal of Consciousness Studies*, vol. 2, no. 4, 1995, pp. 322-326.. URL : <http://ase.tufts.edu/cogstud/papers/unzombie.htm>
- [Dennett 1995-3] Dennett, D. C. 1995. The Practical Requirements for Making a Conscious Robot. URL : <http://www.ecs.soton.ac.uk/~harnad/Papers/Py104/dennett.rob.N@>
- [Dennett 1995-4] Dennett, D. Review on Searle J.R. «The Rediscovery of Consciousness»// Cambridge MA // *Journal of Philosophy*. V. XC. P. 198-199
- [Dennett, 1995-5] Daniel C. Dennett. Cog as a Thought Experiment. FINAL DRAFT for Monte Verità issue, 1995. URL : <http://ase.tufts.edu/cogstud/papers/cogthogt.htm>
- [Dennett, 1999] Dennett Daniel C. (1999), "The zombic hunch: Extinction of an intuition?". URL : <http://www.nyu.edu/gsas/dept/philo/courses/consciousness/papers/DD-zombie.html>
- [Dietrich, 1988] Dietrich, E. 1988. Computers, intentionality, and the new dualism. Manuscript.

- [Dowe, Oppy, 2003] Dowe D., Oppy G. 2003. The Turing Test. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=turing-test>
- [Flanagan, 1991] Flanagan, O. (1991), *The Science of the Mind* (Cambridge, MA: MIT Press), p. 309
- [Flanagan, Polger, 1995] Flanagan Owen & Polger Tom (1995), "Zombies and the function of consciousness", *Journal of Consciousness Studies* 2:313-21.. URL : <http://homepages.uc.edu/~polgertw/Polger-ZombiesJCS.pdf>
- [Flanagan, Polger, 1998] Flanagan, O. & Polger, G. 1998. Consciousness, adaptation, and epiphenomenalism. In (G. Mulhauser, ed) *Evolving Consciousness*.
- [French, 1990] French, R. 1990, Subcognition and the Limits of the Turing Test, *Mind* 99(393), pp. 53–65.
- [French, 1995] French R. (1995), The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It, *Psychology* 7(39). URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
- [French, 1995-1] French, R. M. 1995. Refocusing the Debate on the Turing Test: A Response. *Behavior and Philosophy* 23, pp. 59–60.
- [French, 1995-2] French, R. (1995), The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It, *Psychology* 7(39). <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.. URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
- [French, 1996] French, R. M. 1996, THE INVERTED TURING TEST: HOW A MINDLESS PROGRAM COULD PASS IT (In *Psychology* 7(39) [turing-test.6.french](http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39)). URL : +
- [French, 2000] French, R. M. 2000. The Turing Test: The First Fifty Years In *Trends in Cognitive Sciences*, 4(3), (2000), pp. 115-121.. URL : http://www.ulg.ac.be/cogsci/rfrench/TICS_turing.pdf
- [From ICT to E-culture, 2003] From ICT to E-culture. Advisory report on the digitalisation of culture and the implications for cultural policy. Netherlands Council for Culture, 2003. URL : http://www.cultuur.nl/files/pdf/advies/E-cultuur_engels.pdf
- [Garson, 2002] Garson James, 2002, Connectionism. URL : <http://plato.stanford.edu/entries/connection#@>
- [Genova, 1994-1] Genova, J. Response to Anderson and Keith, *Social Epistemology*, 1994, 8(4), pp. 341–343
- [Genova, 1994-2] Genova, J. Turing's Sexual Guessing Game, *Social Epistemology*, 1994, 8(4), pp. 313–326.
- [Godel, 1931] Godel, Kurt. On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Dover, 1962. URL : <http://www.research.ibm.com/people/h/hirzel/papers/canon00-goedel.pdf>
- [Gulick, 2004] Gulick, Robert Van. Consciousness. Stanford Encyclopedia of Philosophy, 2004. URL : <http://plato.stanford.edu/entries/consciousness/>
- [Gunderson, 1964] Gunderson, K. (1964), *The Imitation Game*, *Mind* 73 pp. 234–245.
- [Harnad, 2001] Harnad, Stevan (2001) *Minds, Machines and Turing: The Indistinguishability of Indistinguishables*. *Journal of Logic, Language, and Information* 9(4):pp. 425-445.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00002615>

- [Harnish, 2001] Harnish, Robert M. *Minds, brains, computers: an historical introduction to the foundations of cognitive science*. Oxford : Blackwell Publishers, 2001. URL : <http://www.worldcat.org/title/minds-brains-computers-an-historical-introduction-to-the-foundations-of-cognitive-science/oclc/223151027>
- [Hauser, 1995] Hauser Larry (1995), "Revenge of the zombies". URL : <http://members.aol.com/lshauser/zombies.html>
- [Hayles, 1999] Hayles, N. Katherine. *How We Became Posthuman*, Chicago: The University of Chicago Press, 1999
- [Hodges, 1997] Hodges, A. (1997). *Turing. A Natural Philosopher*. Phoenix. URL : http://www.amazon.co.uk/exec/obidos/ASIN/0753801922/alanturingwebsite#reader_0753801922
- [Hodges, 2010] Hodges, Andrew (2010). *The Alan Turing Home Page*. URL : <http://www.turing.org.uk/turing/index.html>
- [Hofstadter, 1981] Hofstadter, D. R. 1981. *A coffee-house conversation on the Turing test*. Scientific American.
- [Hopkins, 1998] Hopkins, Patrick D. *Sex/machine: readings in culture, gender, and technology*. Indiana University Press, 1998
- [Horst, 2003] Horst Steven, 2003. *The Computational Theory of Mind*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/computational-mind>
- [Hyslop, 2005] Hyslop, A. 2005. *Other Minds*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/other-minds/>
- [Jefferson, 1949] Jefferson, Geoffrey. *The Mind of Mechanical Man*, *British Medical Journal*, 1(4616):1105–1110, 25 June 1949. URL : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2050428/>
- [Karelis, 1986] Karelis, C. 1986. *Reflections on the Turing Test*, *Journal for the Theory of Social Behavior* 16, pp. 161–172.
- [Kirk, 1974-1] Kirk, R. 1974. *Sentience and behaviour*. *Mind* 81:43-60.
- [Kirk, 1974-2] Kirk, R. 1974. *Zombies vs materialists*. *Aristotelian Society Supplement* 48:135-52.
- [Kirk, 2005] Kirk, R. (2005) *Zombies and Consciousness*. University of Nottingham, November, 2005. URL : <http://www.oup.co.uk/isbn/0-19-928548-9>
- [Kirk, 2006] Kirk, R. 2006. *Zombies*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/zombies/>
- [Kugel, 2002] Kugel Peter. 2002. *Computing Machines Can't Be Intelligent (...And Turing Said So)*. URL : <http://www.cs.bc.edu/~kugel/Publications/Hyper.pdf>
- [Kurzweil, 1999] Kurzweil, Ray. *The Age of Spiritual Machines*, London: Phoenix 1999
- [Lassegue, 1988] Lassegue, J. 1988, *What Kind of Turing Test did Turing Have in Mind?*, *Tekhnema* 3, pp. 37–58.. URL : <http://tekhnama.free.fr/3Lasseguearticle.htm>
- [Lassegue, 1996] Lassegue Jean, 1996. *What Kind of Turing Test Did Turing Have in Mind? Tekhnema 3 / "A Touch of Memory" / Spring 1996*. URL : <http://tekhnama.free.fr/3Lasseguearticle.htm>
- [Leiber, 1989] Leiber, J. 1989. *Shanon on the Turing Test*, *Journal for the Theory of Social Behavior* 19(2), pp. 257–259.
- [Levin, 2004] Levin, J. 2004. *Functionalism*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/functionalism>

- [Lucas, 1961] Lucas, J.R. (1961) Minds, Machines and Goedel. *Philosophy* 36:pp. 112-127.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000356>
- [Lycan, 1987] Lycan, W. G. 1987. Homunctionalism and qualia. In *Consciousness*. MIT Press.. URL : <http://philpapers.org/rec/LYCHAQ>
- [Marton, 1998] Marton, P. 1998. Zombies vs. materialists: The battle over conceivability. *Southwest Philosophy Review* 14:131-38.. URL : <http://www.brown.edu/Departments/Philosophy/zoqvwCTmbie.html>
- [Masrou, 1998] Masrou Farid S. Are Zombies Positively Conceivable? Institute for Studies in Theoretical Physics and Mathematics (IPM). URL : http://www.geocities.com/f_masrur/zombies.html
- [McCarthy, 1969] McCarthy John, 1969, SOME PHILOSOPHICAL PROBLEMS FROM THE STANDPOINT OF ARTIFICIAL INTELLIGENCE. URL : <http://www-formal.stanford.edu/jmc/>
- [McCarthy, 1979] McCarthy, John (1979) Ascribing Mental Qualities to Machines.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000416>
- [McCarthy, 1982] McCarthy John 1982, The Common Business Communication Language. . URL : <http://www-formal.stanford.edu/jmc><http://www-formal.stanford.edu/jmc>
- [McCarthy, 1984] McCarthy, J (1984) Some Expert Systems Need Common Sense.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000423>
- [McCarthy, 1987] McCarthy, John (1987) Generality in Artificial Intelligence; URL : <http://cogprints.ecs.soton.ac.uk/archive/00000425>
- [McCarthy, 1990] McCarthy John, 1990, ARTIFICIAL INTELLIGENCE, LOGIC AND FORMALIZING COMMON SENSE. URL : <http://www-formal.stanford.edu/jmc/>
- [McCarthy, 1995-1] McCarthy, John (1995) Artificial Intelligence and Philosophy.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000420>
- [McCarthy, 1995-2] McCarthy, John (1995) Making Robots Conscious of their Mental States.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000422>
- [McCarthy, 1995-3] McCarthy, John (1995) Todd Moody's Zombies.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000426>
- [McCarthy, 1997] McCarthy, John (1997) The Well-Designed Child.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000428>
- [McCarthy, 1998] McCarthy, John (1998) What is Artificial Intelligence?.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000412>
- [McCarthy, 2001] McCarthy John, 1999, THE ROBOT AND THE BABY. URL : <http://www-formal.stanford.edu/jmc/>
- [Michie, 1974] Michie, D. On Machine Intelligence, New York: Halsted Press, 1974
- [Miłkowski, 2007] Miłkowski, Marcin. Is computationalism trivial?// In: Gordana Dodig Crnkovic and Susan Stuart (eds.), *Computation, Information, Cognition – The Nexus and the Liminal*, Cambridge Scholars Publishing, 2007, pp. 236-246.
- [Millar, 1973] Millar, P.H. (1973), On the Point of the Imitation Game, *Mind* 82, pp. 595–597.
- [Moody, 1995] Moody, Todd (1995), "Conversations with zombies". URL : http://www.imprint.co.uk/Moody_zombies.html

- [Moor, 1976] Moor, J.H. (1976), An Analysis of the Turing Test, *Philosophical Studies* 30, pp. 249–257. URL : <http://www.springerlink.com/content/x336q61282151813/fulltext.pdf>
- [Moor, 1978] Moor, J. H. 1978. Explaining computer behavior. *Philosophical Studies* 34, pp. 325–327
- [Nagel, 1974] Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 4:435–50. Reprinted in *Mortal Questions* (Cambridge University Press, 1979).. URL : <http://hps.elte.hu/~gk/Books/cog-reader/nagel.htm>
- [New Yorker, 1993] On the Internet, Nobody Knows You're a Dog. Page 61 of July 5, 1993 issue of *The New Yorker*, (Vol.69 (LXIX) no. 20). URL : <http://www.unc.edu/depts/jomc/academics/dri/idog.html>
- [O'Hara, Scutt, 1996] O'Hara, K. & Scutt, T. 1996. There is no hard problem of consciousness. *Journal of Consciousness Studies* 3:290–302. Reprinted in Shear 1997.
- [Oppy, Dowe, 2008] Oppy, Graham., Dowe, David (2008). The Turing Test. *Stanford Encyclopedia of Philosophy*. URL : <http://plato.stanford.edu/entries/turing-test/>
- [Owen, 1863] Owen on the Limbs of the Gorilla. Comparison of the Bones of the Limbs of the Troglodytes Gorilla. *Troglodytes niger*, and of Different Varieties of the Human Race; and on the General Character of the Skeleton of the Gorilla by Professor Owen/ *Anthropological Review*, Vol. 1, No. 1 (May, 1863), pp. 149–152
- [Pei, 1995] Wang Pei, 1995. On the Working Definition of Intelligence. URL : <http://www.cogsci.indiana.edu/pub/wang.intelligence.ps>
- [Penrose, 1989] Penrose, R. 1989. *The Emperor's New Mind*. Oxford University Press.
- [Piccinini, 2004] Piccinini, Gualtiero. Functionalism, Computationalism and Mental contents//*Canadian Journal of Philosophy*, Volume 34, Number 3, September 2004, pp.375–410
- [Platt, 1995] Platt, C. 1995, What's It Mean To Be Human, Anyway?, *Wired*.. URL : <http://www.wired.com/wired/archive/3.04/turing.html>
- [Polger, 2004] Polger, Thomas W. Natural minds. *Massachusetts Institute of Technology*: “A Bradford book”, 2004. - 294 p.
- [Polger, 2010] Polger, Tom (web). Computational functionalism. In J. Symons & P. Calvo (eds.), *Routledge Companion to the Philosophy of Psychology*. Routledge. URL : <http://homepages.uc.edu/~polgertw/Polger-CompFunctionalism.pdf>
- [Polger, 2011] Polger Tom, Zombies. URL : <http://host.uniroma3.it/progetti/kant/field/zombies.htm#Top>
- [Povarov, 2001] G. N. Povarov. Semen Nikolayevich Korsakov: Machines for the Comparison of Ideas // G. Trogemann, A.Y. Nitussov , W. Ernst , Eds. *Computing in Russia*. Wiesbaden: Vieweg & Son, 2001
- [Purthill, 1971] Purthill, R.L. (1971), Beating the Imitation Game, *Mind* 80, 290–294
- [Putnam, 1960] 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).
- [Putnam, 1975] Putnam, H. 1975. Philosophy and our mental life. In *Mind, Language, and Reality*. Cambridge University Press.
- [Putnam, 1985] Putnam, Hilary. *Realism and reason*. Cambridge University Press, 1985, 332 p.. URL : <http://books.google.com/books?id=HAjfSA3ir3kC&hl=ru&source=gbssimilarbooks>

- [Raiser, Colby, Faught, Parkinson, 1981] Jon F. Raiser, Kenneth Mark Colby, William S. Faught, and Roger C. Parkinson, "Can Psychiatrists Distinguish a Computer Simulation of Paranoia from the Real Thing? The Limitations of Turing-like Tests as Measures of the Adequacy of Simulations," in *Journal of Psychiatric Research* 15, no. 3 (1980):149-162. Colby discusses PARRY and its implications in "Modeling a Paranoid Mind," in *Behavioral and Brain Sciences* 4, no. 4 (1981): 515-560.
- [Rapoport, 2000] Rapoport William.J. (2000) How to Pass a Turing Test/ *Journal of Logic, Language, and Information* 9: 467-490, 2000. 467. URL : <http://www.cse.buffalo.edu/~rapoport/Papers/TURING.pdf>
- [Ronchi, 2008] Ronchi, Alfredo M. *eCulture. Cultural Content in the Digital Age*, 2008, Approx. 425 p.. URL : <http://www.springer.com/humanities/cultural+studies/book/978-3-540-75273-8>
- [Sampson, 1973] Sampson, G. (1973), In *Defence of Turing*, *Mind* 82, pp. 592–594.
- [Sayers 1941] Sayers, Dorothy L. (1941). *The Mind of the Maker* (1st Ed. ed.). London: Methuen. URL : <http://www.worldinvisible.com/library/dlsayers/mindofmaker/mind.c.htm>
- [Saygin, Cicekli, Akman, 2000] Saygin, A. P. , Cicekli, I. & Akman V. 2000. Turing test: 50 years later. *Minds and Machines* 10:463-518.. URL : <http://www.cs.bilkent.edu.tr/~akman/jour-papers/mam/tt50.ps>
- [Scheffler, 1950] Scheffler, I. 1950. The new dualism: Psychological and physical terms. *Journal of Philosophy*.
- [Schwarz, 2006] Schwarz, Michiel. *e-Culture: Crossovers and Challenges*. Publiziert auf eCulture Factory, 2006. URL : <http://eculturefactory.de/download/schwarz.pdf>
- [Searle, 1980] Searle, J. R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3, pp. 417–424.. URL : <http://www.cs.ucsb.edu/~cs165a/articles/Searle.html>
- [Shanon, 1989] Shanon, B. 1989. A Simple Comment Regarding the Turing Test, *Journal for the Theory of Social Behavior* 19(2), pp. 249–256.
- [Shoemaker, 1975] Shoemaker, S. 1975. Functionalism and qualia. *Philosophical Studies* 27:291-315. Reprinted in *Identity, Cause, and Mind* (Cambridge University Press, 1984).. URL : <http://www.springerlink.com/content/w76571qu83636261/>
- [Sloman, 1978] Sloman, A. 1978. *The Computer Revolution in Philosophy*. Harvester.. URL : <http://www.cs.bham.ac.uk/research/cogaff/crp/>
- [Sloman, Croucher, 1981] Sloman, A. and Croucher, M. (1981) 'Why robots will have emotions.' *Proceedings IJCAI 1981*, Vancouver.. URL : <http://cogprints.org/n@>
- [Sloman, 1996] Sloman, Aaron (1996) What sort of control system is able to have a personality?, in Trapp, Robert and Petta, Paolo, Eds. *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents*, chapter n/a, pages pp. 16-208. Springer.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000702/>>
- [Sloman, 1998] Sloman, Aaron (1998) What Is It Like To Be a Rock?. URL : http://www.cs.bham.ac.uk/~axs/misc/like_to_be_a_rock/rock.html
- [Sloman, 1998] Sloman, Aaron (1998) What Sort of Architecture is Required for a Human-Like Agent?, in *Foundations of Rational Agency*. Kluwer Academic Publishers.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000411/>>
- [Smolensky, 1988] Smolensky, P. 1988. On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11:1-23.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00001675/>

- [Smolensky, 1991] Smolensky, P. 1991. Connectionism, constituency and the language of thought. In (B. Loewer & G. Rey, eds) *Meaning in Mind: Fodor and his Critics*. Blackwell.
- [Smolensky, 1995] Smolensky, P. 1995. On the projectable predicates of connectionist psychology: A case for belief. In (C. Macdonald, ed) *Connectionism: Debates on Psychological Explanation*. Blackwell.
- [Stalker, 1978] Stalker, D. F. 1978. Why machines can't think: A reply to James Moor. *Philosophical Studies* 34, pp. 317–320.
- [Stevenson, 1976] Stevenson, J.G. (1976), On the Imitation Game, *Philosophia* 6, pp. 131–133.
- [Schweizer, 1998] Schweizer, P. 1998. The truly total Turing Test. *Minds and Machines* 8, pp. 263–272.
- [Traiger, 2000] TRAIGER, SAUL. Making the Right Identification in the Turing Test. *Minds and Machines* 10: 561-572, 2000.. URL : <http://www.traiger.net/publications/making-the-right-identification.pdf>
- [Turing, 1936] TURING, A. M. On Coputale Numbers, with an Application to the Entscheidungsproblem. [Received 28 May, 1936.—Read 12 November, 1936.]. URL : http://www.thocp.net/biographies/papers/turing_oncomputablenumbers_1936.pdf
- [Turing, 1950] Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind* 59(236), pp. 433–460.
- [Turing, 1969] Turing, A. (1969), 'Intelligent Machinery', in *ced. Machine Intelligence* 5, Edinburgh University Press, pp. 3–23. Originally, a National Physics Laboratory Report, 1948.
- [van Gulick, 1999] van Gulick, R. 1999. Conceiving beyond our means: The limits of thought experiments. In (S. Hameroff, A. Kaszniak, & D. Chalmers, eds) *Toward a Science of Consciousness III*. MIT Press.. URL : http://cognet.mit.edu/posters/TUCSON3/Van_Gulick.html
- [Watt, 1996] Watt, S. 1996. Naive Psychology and the Inverted Turing Test, *Psychology* 7(14). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000506>
- [Watt, 1998-1] Watt, S. 1998. Seeing this as people: anthropomorphism and common-sense psychology, KMI. URL : <http://citeseer.nj.nec.com/cache/papers/cs/21274/http:zSzzSzkmi.open.ac.ukzSztrSzp aperszSzkmi-tr-61.pdf/watt98seeing.pdf>
- [Watt, 1998-2] Watt, S. 1998. Artificial societies and psychological agents, KMI. URL: <http://citeseer.nj.nec.com/cache/papers/cs/3443/http:zSzzSzkmi.open.ac.ukzSztechrep ortsSzpaperszSzkmi-tr-33.pdf/watt97artificial.pdf>
- [Weizenbaum, 1974] Weizenbaum, Joseph. *CACM* 17, no. 9 (September 1974): 543.
- [Weizenbaum, 1976] Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgement to Calculation*, San Francisco, CA: W.H. Freeman.
- [Weizenbaum, 1996] Weizenbaum, J. (1996), 'ELIZA—A Computer Program for the Study of Natural Language Communication Between Men and Machines', *Communications of the ACM* 9, pp. 36–45.

Alekseev, A. Yu. The Comprehensive Turing Test: the philosophical, methodological and socio-cultural aspects/ A. Alekseev. - Moscow: «AIntell». – 2013. – 304 p.

In the book the first time in the domestic and foreign philosophy of artificial intelligence reveals the idea of the Comprehensive Turing test, which integrates a variety of versions and interpretations of the original Turing test. The Comprehensive test plays a role conceptually categorical interdisciplinary tools of artificial intelligence research, serving as the definition of cognitive phenomena, the principles of designing intelligent computers, the organization of rational discussion on the opportunities and implications of their construction.

The book is addressed to students, undergraduates, interdisciplinary specialists in the field of philosophy, science and engineering of artificial intelligence.

Table of Contents

Preface

Introduction

Chapter 1. Philosophical aspects of the original Turing Test

- 1.1. From Turing machine to the Turing Test: the origins of the constructor's function of the Comprehensive Turing Test
- 1.2. Playing in the simulation of intelligent behavior: the origins of a definite's function of a Comprehensive Turing Test
- 1.3. The polemical standard Turing test: the origins of the critical's functions of a Comprehensive Turing Test

Chapter 2. Private Turing Test

- 2.1. Becoming a Particular Turing Test
- 2.2. Larger versions of the Turing Test
- 2.3 . Ability to pass the Turing Test

Chapter 3. The idea of the Comprehensive Turing Test

- 3.1. Version and interpretation of the Turing Test
- 3.2. Construction of the Comprehensive Turing Test
- 3.3. The formulation of the Comprehensive Turing Test

Chapter 4. The Definite's function of the Comprehensive Turing Test

- 4.1. The definition of social phenomena : the «other» in the e-Culture
- 4.2 . The definition of personal phenomena: the «person» in the e-Culture
- 4.3. The definition of mental phenomena: the «consciousness» in the e-Culture

Chapter 5. The Critical's function of the Comprehensive Turing Test

- 5.1. Criticism of technology: the Comprehensive Turing Test against the complex technology of the Dynamic real-time expert systems
- 5.2. Criticism of the methodology: the Comprehensive Turing Test against a complex project of Artificial Person
- 5.3. Criticism of ideology: the Comprehensive Turing Test against the Complex testing in university education

Chapter 6. The Construct's function of the Comprehensive Turing Test

- 6.1. Construction of philosophical methodology: the justification of the strategic importance of the problem of creativity
- 6.2. Construction of scientific methodology : Neurocomputing - basic technology of Artificial Intelligence
- 6.3. Construction of engineering methodologies: The Integrated Korsakov-Turing machine as fundamental project of the Intelligent computer

Conclusion. The role of Comprehensive Turing Test in methodology of Artificial Societies

Bibliography

Алексеев Андрей Юрьевич

Комплексный тест Тьюринга философско-методологические и социокультурные аспекты

Монография

Утверждено к печати Научным Советом РАН по методологии
искусственного интеллекта, протокол № 50 от «16» декабря 2011 г.

Корректоры: Любовь Стасышина, Екатерина Янковская

Оригинал–макет: Данила Алексеев

Дизайн обложки: Тодор Пожарев

Технический корректор: Василий Алексеев

Издательство «ИИнтелЛ»

109518, г. Москва, ул. Грайвороновская, д. 20 – 171

book@aintell.info, <http://www.aintell.info>,

84991771031

Подписано в печать с оригинал-макета 26 декабря 2011 г.

Формат 60х84 1/16. Ризография. Гарнитура Таймс.

Усл.-печ. л. 16. Уч.-изд.л. 20 Тираж 500 экз.

Заказ № 03 от 20.09.2013 г.

Отпечатано в ЦОП

издательства «ИИнтелЛ»

