

Федеральное государственное бюджетное учреждение науки
Институт философии Российской академии наук

На правах рукописи

Ефимов Альберт Рувимович

**ФИЛОСОФСКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ
ПОСТТЮРИНГОВОЙ ИНТЕЛЛЕКТУАЛЬНОЙ РОБОТОТЕХНИКИ**

Специальность 09.00.08 — философия науки и техники

Диссертация на соискание учёной степени

кандидата философских наук

Научный руководитель:

Дубровский Давид Израилевич — доктор философских наук,
профессор, главный научный сотрудник Института философии РАН

Москва,

2020

Оглавление

Введение	3
Глава 1. Методологический анализ современных исследований	25
в интеллектуальной робототехнике.....	25
1.1. Обобщённый обзор современной интеллектуальной робототехники.....	25
1.2. Основные тренды современной интеллектуальной робототехники в свете методологических проблем создания Общего искусственного интеллекта...	33
1.3. Теоретические и эпистемологические трудности развития интеллектуальной робототехники и Общего искусственного интеллекта	44
Глава 2. Роль тьюринговой методологии в исследованиях интеллектуальной робототехники	70
2.1. Определение значения тьюринговой методологии в контексте интеллектуальной робототехники	70
2.2. Методологическая роль классификации частных тестов Тьюринга в развитии интеллектуальной робототехники	79
2.3. Стена Тьюринга как фундаментальный философско-методологический барьер развития интеллектуальной робототехники	97
Глава 3. Перспективы применения посттьюринговой методологии в интеллектуальной робототехнике.....	110
3.1. Посттьюринговая методология как концептуальный инструментальный разрушения стены Тьюринга.....	110
3.2. Экспериментальный интеллектуальный робот «Э.ЛЕНА» как демонстрация применимости посттьюринговой методологии	122
3.3. Мироззренческие перспективы посттьюринговой коммуникации людей и роботов	133
Заключение	144
Библиография	147

Введение

Актуальность темы исследования. Под интеллектуальной робототехникой понимают совокупность методов, способов исследований и разработки робототехнических систем, интегрирующих возможности робототехники и искусственного интеллекта для создания роботов, интеллектуальных агентов, способных к автономным или частично автономным действиям в различных средах. Междисциплинарный характер интеллектуальной робототехники делает её одним из самых перспективных исследовательских направлений, имеющих значительное влияние на нашу цивилизацию в целом. Этим определяется высокая актуальность разработки философско-методологических вопросов развития интеллектуальной робототехники и её роли в создании *Общего искусственного интеллекта* — стратегической задачи, поставленной сейчас ведущими в технологическом отношении странами, в том числе и в нашей стране. Под Общим искусственным интеллектом понимают программно-аппаратный комплекс, имеющий способность обучаться и действовать лучше человека-специалиста в любой области, достигая поставленных целей в широком диапазоне сред при ограничении доступных ресурсов.

С момента своего возникновения в прошлом веке развитие робототехники и искусственного интеллекта следовало различными траекториями. Робототехника находила применение в промышленности путём адаптации решаемых прикладных задач под ограниченные механические возможности роботов. Искусственный интеллект, оставаясь во многом исследовательской парадигмой, нацеленной на изучение возможностей имитации умственной деятельности человека, использовал широкие возможности программного обеспечения, в частности, методов машинного обучения. Независимость развития этих самостоятельных направлений определялась доступной элементной базой, которая в свою очередь порождала возможности для комбинаторного улучшения.

Робототехника развивалась значительно медленнее в силу присущей ей зависимости от физико-химических свойств и конструкторских особенностей составных частей робота.

Искусственный интеллект, главными компонентами которого являются математическое и программное обеспечение, развивался значительно быстрее робототехники в силу прямой зависимости от исследований и разработок в области микроэлектроники, до сих пор следующих эмпирическому закону Мура¹ (количество транзисторов, размещаемых на кристалле интегральной схемы, удваивается каждые 24 месяца). Это дало искусственному интеллекту возможность использовать экспоненциальное развитие элементной базы компонентов, таких как графические ускорители, системы хранения данных, различные сенсоры и широкополосные связи для создания программного обеспечения, развивающего подходы машинного обучения к решению большого типа прикладных задач.

Результатом этого явилась ситуация, в которой методы искусственного интеллекта используются для решения «куновских» головоломок робототехники, ускоряя прогресс обоих направлений. Новые средства и среды программирования, такие как искусственные нейронные сети глубокого обучения, позволяют находить ответы на сложные задачи робототехники. В свою очередь, это порождает новые способы человеко-машинного взаимодействия, требующие глубокого философско-мировоззренческого и эпистемологического осмысления с привлечением принципов постнеклассической рациональности (В. С. Степин, В. А. Лекторский и др.). Именно практическая востребованность новых методологических подходов определяет важную роль участия философов в современных разработках искусственного интеллекта.

¹ *Leiserson C. E. et al.* There's plenty of room at the Top: What will drive computer performance after Moore's law? // *Science* 368 (6495), eaam9744. URL: <https://science.sciencemag.org/content/368/6495/eaam9744> (дата обращения: 28.07.2020).

Развитие технологий искусственного интеллекта и робототехники подошло к рубежу, когда их влияние на нашу цивилизацию выходит за пределы экономической сферы и становится важным фактором геополитической и экологической глобальной повестки, наряду с тем, что продолжает оставаться одной из центральных проблем научно-технического развития. Всемирный экономический форум² (WEF, 2019), признавая стратегический приоритет этих технологий, включает их в число «технологий слияния» (fusion technologies). Результаты развития искусственного интеллекта и робототехники могут быть источником новых рисков и системных проблем (нарушение неприкосновенности частной жизни, изменение самого понятия занятости и др.), вместе с этим они обещают успешный рост экономики, парирование ряда негативных следствий нарастающего глобального кризиса нашей цивилизации.

Вместо стимулирования кооперации усиливается международная конкуренция за первенство в создании Общего искусственного интеллекта. Президент России В. В. Путин подчеркивает, что искусственный интеллект «действительно одно из ключевых направлений технологического развития, которые определяют и будут определять будущее всего мира»³.

Президент США Д. Трамп проводит политику своей страны на сохранении лидерства в искусственном интеллекте, так как «это имеет критическое значение для поддержания экономической и национальной безопасности Соединённых Штатов»⁴.

Всё это превращает искусственный интеллект и робототехнику в геополитические инструменты, делая технологии заложниками интересов

² Сайт Всемирного экономического форума в Давосе. URL: <https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2?tab=publications> (дата обращения: 28.07.2020).

³ Совещание по вопросам развития технологий в области искусственного интеллекта. URL: <http://kremlin.ru/events/president/news/60630> (дата обращения: 28.07.2020).

⁴ Artificial Intelligence for the American People. URL: <https://www.whitehouse.gov/ai/> (дата обращения: 28.07.2020).

транснациональных корпораций. Создается некое подобие «корпоративного» «Манхэттенского проекта»⁵.

Искусственный интеллект и робототехника, наряду с энергетикой, транспортом, являются технологией, которая, с одной стороны, выступает как самостоятельная отрасль промышленного производства, а с другой — создает производительные силы для остальных отраслей. Такую технологию обычно называют технологией общего назначения или сквозной технологией. Благодаря проникновению во все сферы социальной жизни робототехника создает эффект, сходный с эффектом изобретения паровой машины в эпоху первой промышленной революции⁶.

Необходимо отметить, что для Общего искусственного интеллекта и робототехники сложились необходимые условия: появляются новые инструменты (материалы, программные архитектуры и пр.), а также новые виды деятельности, позволяющие этот замысел реализовать. Это справедливо подчеркивается В. М. Розиным⁷, который писал, что робототехника оказалась в зоне «ближайшего технологического развития» в силу указанных выше причин.

Робототехника и искусственный интеллект могут также быть и предметом большой опасности. Машины создают условия для неразличимости действительных фактов и лжи — искусственные нейронные сети стали основой для генерации псевдореальности, так называемых *deepfakes*, которые способствуют глубокому кризису экспертизы, расцвету релятивизма и скепсиса не только в обыденном знании, но и в знаниях,

⁵ *Levin J.C. et al. Roadmap to a Roadmap: How Could We Tell When AGI is a 'Manhattan Project' Away? // ECAI 2020. URL: <https://arxiv.org/abs/2008.04701> (дата обращения: 20.09.2020).*

⁶ *Crafts N. Steam as a GPT: A growth accounting perspective // The Economic journal. 2004. No. 114. P. 338–351.*

⁷ *Розин В. М. Эволюция инженерной и проектной мысли. Инженерия: Становление, Развитие, Типология. М.: Ленанд, 2015. С. 123.*

создаваемых в различных исследовательских центрах (В. Е. Лепский)⁸. Поэтому необходимо пристальное изучение в оптике философии и методологии всех особенностей развития искусственного интеллекта как отрасли науки, производства и образования. Ряд аспектов, таких как персонализация образовательных траекторий, автоматическая генерация учебных курсов и т. п. нуждаются в тщательном и всестороннем обсуждении (А. Л. Семенов)⁹.

В последние десятилетия, несмотря на успешное развитие науки и технологий в области искусственного интеллекта, у исследователей и инженеров нет ясного ответа на вопрос, как развивать технологии в горизонте следующих десятилетий. Причина в том, что теоретико-методологическая и философская основа такого планирования разработана весьма слабо. Целиком сохраняется актуальность мысли К. Ясперса, что «вся дальнейшая судьба человека зависит от того способа, посредством которого он подчинит себе последствия технического развития и их влияние на его жизнь»¹⁰.

В свое время Т. Кун (научные парадигмы) и И. Лакатос (исследовательские программы) создали основу планирования в естественно-научных исследованиях, основываясь на развитии физических наук в XIX и первой половине XX века. Однако главные открытия в области искусственного интеллекта и робототехники приходится на вторую половину XX и начала XXI веков. Философско-методологические основы науки, которые были разработаны на основе естественно-научных прорывов XX века

⁸ Лепский В. Е. Асимметричный ответ информационным войнам XXI века // Рефлексивные процессы и управление: Сборник материалов XI Международного симпозиума (16–17 октября 2017 г., Москва) / Отв. ред. В. Е. Лепский. М.: Когито-Центр, 2017. С. 221–224.

⁹ Семенов А. Л. Революция искусственного интеллекта и общее образование: доклад на 99-ом заседании семинара «Философско-методологические и научно-теоретические проблемы ИИ» Научного совета по методологии искусственного интеллекта при Отделении общественных наук РАН (НСМИИ РАН), г. Москва, Институт философии РАН, 16 января 2020 г. URL: <https://drive.google.com/file/d/1aRuT4a7qGICzmsJhS1EZPPYp5hhCOEis/view> (дата обращения: 20.08.2020).

¹⁰ Ясперс К. Современная техника / Пер. М.И. Левиной // Новая технократическая волна на Западе / Сост. П.С. Гуревич. М.: Прогресс, 1986. С. 119–146.

и стали основой для развития космонавтики и атомной энергетики, не могут служить надёжным ориентиром в развитии нынешнего этапа искусственного интеллекта и интеллектуальной робототехники.

Наблюдаются постоянные расхождения в прогнозах создания Общего искусственного интеллекта. Вычислительные мощности, которые казались А. Тьюрингу достаточными для создания «мыслящей машины», давно достигнуты; возможности создания, хранения и передачи информации в интернете превышают самые смелые мечты полувековой давности. Однако все это лишь частные результаты, далёкие от создания Общего интеллекта.

Такое положение дел обращает нас к внимательному рассмотрению методологии, разработанной А. Тьюрингом в 1940–50-е годы, к анализу исторических ограниченностей этой методологии, которые должны быть преодолены на нынешнем, новом этапе развития искусственного интеллекта, когда ставится задача создания Общего искусственного интеллекта. Эти ограниченности, недостатки методологии А. Тьюринга можно представить следующим образом: а) узкая интерпретация парадигмы функционализма в форме сугубо бихевиористской и операционалистской трактовки интеллекта. С этим связаны: б) недостаточность определения критериев разумного поведения машины (тест Тьюринга); в) жёсткое разделение человека и машины при изучении их взаимодействия («стена Тьюринга»); г) стремление создать «мыслящую программу», но не «мыслящего робота». К этому можно добавить ряд других существенных пунктов, связанных с игнорированием проблемы сознания, преувеличением роли символических систем в интеллектуальной деятельности и др. Эти ограниченности и недостатки должны быть преодолены в рамках *посттьюринговой методологии*, которая призвана к созданию новых теоретических установок, подходов и методов для разработки специфических когнитивных архитектур Общего искусственного интеллекта (разумеется, нисколько не принижая при этом выдающихся эпохальных достижений А. Тьюринга).

В работах по методологии науки важное место всегда занимало осмысление парадигмы и программы новых научных исследований. В ходе этих исследований при обсуждении проблематики искусственного интеллекта не раз возникали предложения опоры на так называемые «конвергентные технологии», которые включают в себя нано-технологии, био-технологии, информационные технологии и когнитивные технологии (обозначаемые аббревиатурой NBIC), которые должны были стимулировать междисциплинарный подход к развитию искусственного интеллекта и робототехники¹¹. Однако это не оправдало себя в достаточной мере (И. Ю. Алексеева, В. И. Аршинов)¹², поскольку масштабные исследовательские программы фокусировались на первой части парадигмы конвергентных технологий (нано-, био-), отодвигая вторую часть (инфо-, когно-) на более далекую перспективу.

Интеллектуальная робототехника выступает одновременно в двух качествах. С одной стороны, это область научных исследований, посвящённая изучению принципов, законов взаимодействия физических объектов, наделённых «интенциональностью», переданной им человеком, и способных передавать человеку информацию о результатах выполнения поставленной задачи. С другой стороны, робототехника включает такие разнородные по своей природе группы технологий как роботы-исполнители хозяйственных функций, роботы-автомобили, дроны, всевозможные промышленные манипуляторы и т. п.

Можно сказать, что робототехника, как раздел техники, занимает положение «моста», связки между быстро растущими возможностями искусственного интеллекта в виртуальном мире и относительно статичной инфраструктурой, созданной цивилизацией (города, дороги, производство,

¹¹ Аршинов В. И. На пути к антропному измерению NBIC-конвергенции // Философские проблемы биологии и медицины. Вып. 4. М., 2010. С. 41–44.

¹² Алексеева И. Ю., Аршинов В. И. Информационное общество и НБИКС-революция. М.: ИФ РАН, 2016. 196 с.

системы коммуникации и пр.). Искусственный интеллект наделяет любое техническое устройство возможностью изменять своё поведение на основе данных, поступающих от сенсоров, тем самым превращая его в робота: открываются новые возможности интеграции человека и робота, о которых говорил Ю. Хабермас¹³.

По мере совершенствования технологий робототехники и искусственного интеллекта, эмпирическое определение человека также меняется. Киборгизация, виртуальные интеллектуальные ассистенты, устройства, исполняющие роль экзокортекса, внутренних и внешних органов, телеприсутствие — всё это искусственные дополнения (расширения) возможностей человека, без которых наш вид в течение нескольких следующих десятилетий не сможет мыслить своё существование¹⁴. Однако эти «дополнения» ставят вопросы об определении того, что составляет саму сущность человека и как отличить человека от машины, которая способна самостоятельно выполнять действия, ранее считавшиеся исключительно «человеческими». Можно ли принять результат экзамена молодого человека, которому оказал помощь его личный цифровой ассистент? Если нельзя, то в чём принципиальная разница между цифровым ассистентом и калькулятором? Другими словами, идёт встречный процесс, меняется не только «телесность» роботов, которая становится всё более разнообразной, в том числе и антропоморфной, но и телесность человека, которая всё больше технологизируется. Это делает невозможным использование классического рационального подхода, в рамках которого люди это субъекты, а машины - объекты. Исследователи искусственного интеллекта существуют в рамках саморазвивающейся системы, которой является отрасль высоких технологий. Поэтому здесь необходимо использование принципов постнеклассической

¹³ Хабермас Ю. Будущее человеческой природы. М.: Весь Мир, 2002. 144 с.

¹⁴ Лекторский В. А. Совместимы ли наука и утопия? // Мифология века НТР. Утопии, мифы, надежды и реальность новейших направлений науки. От Франкенштейна и эликсира бессмертия до «биокиборгов» и постчеловека / Отв. ред. Г. А. Белкина. М.: URSS, 2019. С. 35–49.

рациональности, как эпистемологической основы исследования и обоснования: посттьюринговой методологии развития искусственного интеллекта и интеллектуальной робототехники.

На раннем этапе развития искусственного интеллекта и робототехники (в работах А. Тьюринга, Дж. фон Неймана, Н. Винера и др.), преобладал узкодисциплинарный подход, сочетающий в себе сравнительно ясные методологические и этические аспекты и практические шаги по развитию технологий (машина Тьюринга, архитектура фон Неймана). Однако современная специализация искусственного интеллекта и его новейшие результаты обнаруживают всё же существенный разрыв между теоретическими исследованиями в области философии и методологии искусственного интеллекта с одной стороны, и практическими усилиями исследователей по созданию интеллектуальных машин (роботов), которые будут обладать способностями, не уступающими человеку — с другой. Подобный разрыв делает исследования в области методологического подхода к проблемам искусственного интеллекта и робототехники крайне актуальными и требует разработки специальных междисциплинарных программ и возникающих при этом острых эпистемологических вопросов (В. Г. Буданов)¹⁵.

Инженерно-технические проблемы (куновские «головоломки») имеют свои решения и решаются многочисленными коллективами в крупных научных центрах и предприятиях по всему миру. Но проблема долгосрочного развития Общего искусственного интеллекта и интеллектуальных роботов имеет междисциплинарный характер в силу глобального значения самого замысла этих технологий. Идея опоры на «закон Мура», на чисто количественное наращивание вычислительных мощностей при сохранении концептуального подхода А. Тьюринга — это методологический тупик. Новые актуальные подходы и методологии могут не только сократить время и

¹⁵ Штукельбергер К., Буданов В. Г., Олескин А. В. и др. Вызовы будущего: искусственный интеллект, технологии, этика // Экономические стратегии. 2019. Т. 21. № 6 (164). С. 18–29.

силы на создание нового поколения умных и полезных машин, но и реализовать потенциал искусственного интеллекта и робототехники для решения актуальных цивилизационных вызовов.

Степень научной разработанности проблемы. Философско-методологические основы разработки искусственного интеллекта и робототехники всегда привлекали внимание ведущих исследователей. Исходная позиция была задана с самого начала А. Тьюрингом в рамках классической рациональности, сфокусированной на возможностях «мыслящих машин», как объектов исследования. В своей работе 1950 г. «Вычислительные машины и разум» британский математик и логик заложил основы операционалистического, анти-эссенциалистского подхода к созданию «мыслящих машин» ещё до того, как возник сам термин «искусственный интеллект».

За 70 с лишним лет, прошедших с момента этой публикации, были выдвинуты различные подходы к определению философско-методологических основ робототехники. Начиная с 1956 г., когда Дж. Маккарти впервые употребил термин «искусственный интеллект», эта область связывалась с математическими, лингвистическими и алгоритмическими проблемами, необходимыми для имитации интеллекта человека с помощью компьютера.

Параллельно с философско-методологическими обоснованиями искусственного интеллекта развивается предметная область «интеллектуальной робототехники», под которой имеют в виду робототехнику, способную решать сложные интеллектуальные задачи. Термин «робототехника» («robotics») появился в 1941 г. благодаря А. Азимову, хотя само слово «робот» — ещё в 1920 г. у чешского писателя К. Чапека¹⁶.

¹⁶ В пьесе К. Чапека «RUR» роботами были искусственно созданные человекоподобные биологические машины.

С момента окончания Второй мировой войны, которая дала значительный импульс развитию вычислительной техники, использовавшейся для криптоанализа и баллистических вычислений, возникла неразрывная связь между искусственным интеллектом и робототехникой (Дж. Маккарти и Дж. Энгельбергер)¹⁷. Первое направление рассматривалось исследователями широко как теоретическая возможность имитации интеллекта человека на небиологическом субстрате, второе направление всегда ставило задачей реальную имитацию действий человека роботом в физическом мире.

В своих основополагающих работах по искусственному интеллекту А. Тьюринг подчеркивал связь между когнитивными функциями мозга и моторикой человека. Это послужило основой другой теоретико-практической дисциплины — кибернетики, которую основал Н. Винер.

Термин «интеллектуальная робототехника» появился в 60–70-х годах прошлого века благодаря работам двух исследовательских групп в США. В Стенфордском исследовательском институте был создан автономный робот «Shakey», способный полностью к передвижению в физическом пространстве. Другим направлением было создание Т. Виноградом в MIT примерно в то же время программы «SHRDLU», понимающей ограниченное подмножество естественного языка и действующей в виртуальном мире простых геометрических тел — блоков. Эту программу можно было назвать «виртуальным роботом».

Принципы развития искусственного интеллекта в конце XX века как области создания программного обеспечения, которая может помочь сделать компьютеры «интеллектуальнее», была подвергнута значительному переосмыслению в силу разнообразных причин, включая завышенные ожидания результатов исследований в этой области (М. Мински). В 90-х годах прошлого века стали раздаваться первые призывы к объединению

¹⁷ *Markoff J. Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots. New York: Harper Collins Publishers, 2015. 400 p.*

направлений «искусственный интеллект» и «робототехника» в единое понятийное пространство (Р. Брукс)¹⁸.

Другой причиной появления термина «интеллектуальная робототехника» являлась постоянная критика направления, обозначаемого термином «искусственный интеллект», как якобы слишком далекого от реальных достижений (Р. Пенроуз)¹⁹. В попытках дифференцировать свои исследования многие учёные прибегают к созданию новых терминов, таких как «интеллектуальные системы», «когнитивные системы» или «когнитивная робототехника». Однако нет никакой необходимости в разборе тонкостей терминологии, так как «интеллектуальная робототехника» по сути полностью закрывает все возможности применения методов искусственного интеллекта к робототехнике, являясь вполне адекватным общим термином (Р. Мёрфи)²⁰. Таким образом, интеллектуальная робототехника является операционалистским направлением, основанным большей частью на работах Х. Моравека²¹, Д. Ликлидера²², определяющим предметную область через действия агентов, которые должны отвечать функциональному описанию трёх одновременно реализуемых способностей: **ощущения, понимания и действия**.

Философское осмысление вопросов робототехники и искусственного интеллекта проводилось в работах таких представителей аналитической философии как Х. Патнэм, Дж. Фодор, Т. Нагель, Дж. Сёрл, Д. Деннет, Д. Чалмерс, Н. Блок и др.

¹⁸ *Brooks R. A. Intelligence Without Representation // Artificial Intelligence. 1991. No. 47. P. 139–159.*

¹⁹ *Пенроуз Р. Новый ум короля: о компьютерах, мышлении и законах физики / Пер с англ. В. О. Малышенко. М.: УРСС, ЛКИ, 2011. 400 с.*

²⁰ *Murphy R. Introduction to AI Robotics. Cambridge: The MIT Press, 2019. 648 p.*

²¹ *Moravec H. Mind Children: The Future of Robot and Human Intelligence. Cambridge, MA: Harvard University Press, 1988. 224 p.*

²² *Licklider J. C. R. Man-Computer Symbiosis // IRE Transactions on Human Factors in Electronics. 1960. Vol. HFE-1. P. 4–11.*

Следует особо отметить значительный вклад в разработку этих вопросов, внесённый советскими философами ещё в 60–80-х годах прошлого века (Б. Г. Бирюков, А. Г. Спиркин, И. Б. Новик, А. Д. Урсул, Л. Б. Баженов, В. С. Тюхтин, П. В. Копнин, Л. А. Петрушенко и др.). В отличие от западных философов они более широко ставили эти вопросы в плане задач кибернетики, разработки концепции информации и развития самоорганизующихся систем, включая проблему сознания. В этих исследованиях принимали активное участие крупнейшие советские ученые: А. И. Берг, А. Н. Колмогоров, Н. А. Бернштейн, П. К. Анохин, В. М. Глушков и др., вклад которых трудно переоценить. Вся эта продуктивная и многоплановая работа советских философов и учёных требует специального исследования.

В последние годы актуальные вопросы развития искусственного интеллекта широко обсуждались на конференциях высокого уровня. Среди них можно выделить: контексты компьютерной модели «Я», включение конфликтологии и принципа ограниченной рациональности в изучение многоагентных систем (М. А. Шестакова²³), проблему эмоциональных реакций и телесности виртуальных людей (Д. Бурден, М. Савин-Баден²⁴), проблемы социализации роботов, их включение в публичные пространства (О. Шерер, Ж. Парвиенен²⁵), проблемы подходов различных культур в создании роботов, попытки осмыслить феномен «души» компьютеров (К. Джеймс²⁶). Разумеется, широко обсуждаются вопросы этики и безопасности использования роботов.

²³ Шестакова М. А. Методологические проблемы многоагентных систем // Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвящённой шестидесятилетию исследований искусственного интеллекта, 17–18 марта 2016 г. М.: ИИнтелл, 2017. С. 189–202.

²⁴ Burden D., Savin-Baden M. *Virtual Humans. Today and Tomorrow*. Boca Raton: CRC Press, 2019. 270 p.

²⁵ *Envisioning Robots in Society — Power, Politics, and Public Space Proceedings of Robophilosophy 2018 / TRANSOR-2018 (February 14–17, 2018, University of Vienna, Austria)* / Ed. by M. Coeckelbergh, J. Loh, M. Funk, J. Seibt and M. Nørskov. Amsterdam: IOS PRESS BG, 2018. 391 p.

²⁶ James K. *The Battle for the Robot Soul* // *Philosophy Now*. 2020. No. 139. P. 16–19.

Необходимо отметить, что исследователи-робототехники по вполне понятным причинам сосредоточены на решении сугубо прикладных задач, фокусируют внимание в основном на частных, ограниченных примерах взаимодействия робота и окружающей среды, оставляя большей частью в стороне подходы, которые требуют решения теоретических и философско-методологических вопросов общего порядка. Разработка этих вопросов, касающихся специально интеллектуальной робототехники, пока ещё слабо представлена в нашей философской литературе, хотя они являются важным условием её дальнейшего эффективного развития, особенно когда речь идёт о прорывных достижениях в этой области.

Соответственно, современные разработчики искусственного интеллекта стремятся к некоему идеалу, задаваемому действиями агентов. Такой идеал описывает формулировка Б. Герцля: «Общий искусственный интеллект — это способность обучаться и действовать лучше специалиста в любой области, достигая целей в широком диапазоне сред при ограничении ресурсов»²⁷.

Важным стимулом для формирования посттьюринговой методологии интеллектуальной робототехники явились труды А. Ю. Алексева²⁸ в области комплексного теста Тьюринга.

Диссертация опирается на результаты исследований в области философских проблем искусственного интеллекта и смежных с ними вопросов (особенно эпистемологического характера), проведённых отечественными философами (работы И. Ю. Алексеевой, А. М. Анисова, В. И. Аршинова, В. Г. Буданова, В. Л. Васюкова, И. А. Герасимовой, В. А. Глазунова, Д. И. Дубровского, И. Т. Касавина, Л. П. Киященко, Е. Н. Князевой, В. А. Лекторского, В. Е. Лепского, И. К. Лисеева, В. М. Розина, В. С. Степина, Е. О. Труфановой, В. К. Финна, В. И. Шалака, Е. Н. Шульги и др.); в

²⁷ *Goertzel B. Artificial general intelligence: Concept, state of the art, and future prospects // Journal of Artificial General Intelligence. 2014. Vol. 5. P. 1–46.*

²⁸ *Алексеев А. Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга: диссертация на соискание ученой степени доктора философских наук: 09.00.08. М., 2015. 482 с.*

исследовании обсуждаются и предлагаются решения актуальных теоретико-методологических и философских вопросов нынешнего этапа развития интеллектуальной робототехники, создания принципиально новых когнитивных архитектур, отвечающих задачам построения Общего искусственного интеллекта.

Объектом исследования выступают философско-методологические концепции и подходы интеллектуальной робототехники, отражающие изменения основ создания базовых и прикладных технологий в этой области: появления больших массивов семантически размеченной информации для машинного обучения и специализированных ускорителей вычислений, облегчающих обучение на этих данных. Программа исследований, которая была инициирована А. Тьюрингом и брала за основу сугубо операционалистский подход, в значительной степени устарела, так как симуляция интеллектуального поведения человека возможна с помощью искусственных нейронных сетей глубокого обучения. Поэтому необходима разработка новых подходов в исследованиях искусственного интеллекта.

Предметом исследования являются концептуальные основы (принципы) интеллектуальной робототехники в свете посттьюринговой методологии.

Материалом для диссертационного исследования являются разнообразные междисциплинарные теоретические, эмпирические и философские работы как зарубежных, так и российских исследователей, включая собственные технические исследования автора.

Источники диссертации

В диссертационном исследовании были использованы научные труды и публикации философского, естественнонаучного и прикладного характера на русском и английском языках.

Цель и задачи диссертационного исследования

Цель диссертационного исследования состоит в комплексном анализе основных технологических трендов, современного состояния и научных перспектив *интеллектуальной робототехники*, способов преодоления её теоретических и методических трудностей на основе их философско-методологического осмысления и разработки предлагаемой автором концепции посттьюринговой методологии.

Общая цель конкретизируется в следующих *задачах*:

- проанализировать современное состояние интеллектуальной робототехники и приоритетные направления исследований в этой области;
- провести классификацию частных тестов Тьюринга, уточнить их роль в развитии интеллектуальной робототехники, показать историческую ограниченность методологии Тьюринга в свете современных задач интеллектуальной робототехники;
- рассмотреть основные концептуальные барьеры в развитии интеллектуальной робототехники и предпосылки для их преодоления; показать, что для этой цели необходимо учитывать результаты философских и феноменологических разработок проблемы сознания (которые игнорирует классическая парадигма искусственного интеллекта);
- описать феномен «стены Тьюринга» и обосновать принципы посттьюринговой методологии (разрушающей эту «стену»), показать её преимущества в сравнении с классической тьюринговой методологией;
- продемонстрировать возможности использования созданного под руководством диссертанта робота «Э.ЛЕНА» как экспериментального инструмента для анализа сферы применимости посттьюринговой методологии;
- наметить основные мировоззренческие перспективы посттьюринговой коммуникации людей и роботов.

Методологической основой исследования выступает, в первую очередь, развитая в российской философии концепция постнеклассической эпистемологии в её отношении к классической и постклассической эпистемологии (В. А. Лекторский, В. С. Степин, В. Г. Буданов, В. И. Аршинов и др.), позволяющая глубже осмыслить особенности познавательных и деятельных возможностей субъекта (человека), влияние на него объекта (робота) и средства наблюдения (тест Тьюринга) в условиях междисциплинарного развития науки и технологий и спроецировать их на взаимоотношения человека и робота, взаимодействие которых разделено средством наблюдения («стена Тьюринга»). Процессы развития искусственного интеллекта и интеллектуальной робототехники осмыслены в рамках постнеклассической рациональности, что позволяет наметить эффективные пути их дальнейших исследований в посттьюринговой методологии.

Научная новизна диссертационной работы определяется созданием концептуального подхода, теоретически обосновывающего переход к новой посттьюринговой методологии как основы для прорыва в интеллектуальной робототехнике и в разработках Общего искусственного интеллекта.

Научная новизна диссертационного исследования может быть сформулирована следующим образом:

- проведён анализ текущего состояния, перспектив и барьеров развития интеллектуальной робототехники в контексте тьюринговой методологии, который показал, что фундаментальные технологические барьеры могут быть пройдены, в силу чего открываются возможности для создания интеллектуальных роботов нового уровня, которые будут способствовать созданию Общего искусственного интеллекта;
- проанализированы и систематизированы частные тесты Тьюринга, выявлены их концептуальные возможности в двух измерениях: *виртуальном-физическом* и *невербальном-вербальном*, что даёт

основание для введения понятия «техно-умвельта», определяющего мир робота и содержащего срез отображения мира, каким его «видит» робот (искусственный интеллект), а также набор действий, доступных для данного интеллектуального робота. Также выделены и описаны четыре «техно-умвельта», которые охватывают весь мыслимый диапазон действий робота;

- предложен новый подход для развития научных исследований в области интеллектуальной робототехники на основе посттюринговой методологии, которая обосновывает способы создания интеллектуальных роботов, обладающих возможностями самостоятельного действия одновременно в плане различных измерений — *виртуального-физического* и *невербального-вербального*;
- уточнено понятие Общего искусственного интеллекта в применении к интеллектуальным роботам, действующим в различных «техно-умвельтах»;
- рассмотрены социально-культурные и мировоззренческие перспективы производственной и бытовой коммуникации людей и роботов.

Теоретическая и практическая значимость исследования.

Выполненное в данной диссертационной работе исследование позволило теоретически обосновать посттюринговую методологию для развития интеллектуальной робототехники и создать методологическую основу для классификации частных тестов Тьюринга. Исследование позволяет по-новому подойти к разработке когнитивных архитектур для технологий интеллектуальной робототехники и Общего искусственного интеллекта на основе переходов из различных доменов тьюрингового пространства. Практические результаты, включая опубликованные статьи и полученный автором патент, служат расширению деятельных возможностей «воплощённого интеллекта» и его способностей невербального общения в физическом и виртуальном мире при взаимодействии с человеком.

Положения, выносимые на защиту:

- 1 Введён новый способ классификации частных тестов Тьюринга, в соответствии с тем, насколько тесты используют *виртуальное-физическое* или *вербальное-невербальное* взаимодействие, что позволяет систематизировать подход к их изучению и их применению для развития Общего искусственного интеллекта и интеллектуальной робототехники. Данная классификация позволяет обосновать понятие «техно-умвельта» и даёт возможность рассмотреть методологию А. Тьюринга в оптике постнеклассической научной рациональности и показать её историческую ограниченность.
- 2 Выявлена роль «стены Тьюринга» в тьюринговой классической методологии как барьера в осуществлении продуктивного взаимодействия человека и компьютера. Устранение «стены Тьюринга» позволяет перейти от классического сравнения действий человека и машины к эффективной их коллаборации в смысле постнеклассической парадигмы науки (для решения задач создания Общего искусственного интеллекта).
- 3 Введено понятие посттьюринговой методологии и разработаны её основные положения, которые могут служить созданию новых когнитивных архитектур Общего искусственного интеллекта и интеллектуальной робототехники, так как обеспечивают решение задачи трансляции опыта и знаний интеллектуального робота из различных «техно-умвельтов».
- 4 Рассмотрены на сценарных примерах вероятные мировоззренческие перспективы посттьюринговой коммуникации людей и роботов в области образования, производства и развлечения.

Достоверность научных положений, выводов и рекомендаций подтверждается:

- адекватностью применяемых научно-методологических подходов к цели и задачам исследования;

- проведением научного исследования с опорой на междисциплинарный анализ;
- согласованностью выводов с существующими актуальными исследованиями в области искусственного интеллекта, робототехники, постнеклассической эпистемологии и синергетики;
- апробацией результатов исследования.

Апробация материалов исследования

Основное содержание диссертационного исследования отражено в четырёх статьях, входящих в Перечень ведущих рецензируемых журналов и изданий, рекомендуемых ВАК Минобрнауки РФ, а также входящих в международные реферативные базы данных и системы цитирования (Scopus и Web of Science), и 2-х патентах:

- Ефимов А. Р. Будущее искусственного интеллекта: тьюринговая или посттьюринговая методология? // Искусственные общества. 2019. Т. 14. № 4. DOI: 10.18254/S207751800007698-6. URL: <https://artsoc.jes.su/s207751800007698-6-1/> (в соавт. с А. Ю. Алексеевым, В. К. Финном);
- Ефимов А. Р. Технологические предпосылки неразличимости человека и его компьютерной имитации // Искусственные общества. 2019. Т. 14. № 4. DOI: 10.18254/S207751800007645-8. URL: <https://artsoc.jes.su/s207751800007645-8-1/>;
- Ефимов А. Р. Посттьюринговая методология: разрушение стены на пути к общему искусственному интеллекту // Интеллект. Инновации. Инвестиции. 2020. № 2. С. 74–80. DOI: 10.25198/2077-7175-2020-2-74. URL: http://intellekt-izdanie.osu.ru/arch/2020_2_74.pdf;
- Efimov A. Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // Artificial General Intelligence (AGI). 2020. Vol. 12177. P. 83-94. DOI: 10.1007/978-3-030-52152-3_9.

- Ефимов А. Р. Способ и система для создания мимики на основе текста // Патент на изобретение RU 2 723 454. 2020 (в соавт. с А. С. Гонноченко, М. А. Владимировым);
- Ефимов А. Р. Способ и система предиктивного избегания столкновения манипулятора с человеком // Патент на изобретение RU 2 685 996 (в соавт. с А. Л. Постниковым, А. Р. Гамаюновым, Д. Д. Зятыговым).

Основные результаты диссертационного исследования обсуждались на заседаниях сектора междисциплинарных проблем научно-технического развития Института философии Российской академии наук и в форме докладов были представлены на следующих международных и всероссийских конференциях:

- 4-й Международной конференции «Больше, чем обучение: как добиться изменения поведения?» (г. Москва, 25 октября 2019);
- Всероссийской научной конференции «Суперкомпьютерные технологии в общественных науках» (г. Москва, 24 сентября 2019);
- Международного междисциплинарного семинара «Искусственный интеллект: методология и технология» (г. Москва, 20 сентября 2019);
- Международной конференции AI Journey (г. Москва, 08 ноября 2019). На этой конференции выступал Президент России В. В. Путин;
- Международной междисциплинарной конференции «Философия искусственного интеллекта» (г. Москва, июнь 2020).

Кроме того, по теме робототехника и искусственный интеллект выполнены другие публикации и исследовательские работы:

- Ефимов А. Р. Снятся ли чат-ботам андройды? Перспективы технологического развития искусственного интеллекта и робототехники // Философские науки. 2020. Т. 62. № 7. С. 73-95. DOI: 10.30727/0235-1188-2019-62-7-73-95.
URL: https://www.phisci.info/jour/article/view/2697?locale=ru_RU;
- Ефимов А. Р. и др. Практическое применение роботов и сопутствующих технологий в борьбе с пандемией COVID-19 // Робототехника и

- техническая кибернетика. 2020. Т. 8. № 2. С. 87–100. DOI: 10.31776/RTCJ.8201. URL: <https://rusrobotics.ru/index.php/tom-8-nomer-2-2020/273-obshchie-voprosy/719-prakticheskoe-primeneniye-robotov-i-soputstvuyushchikh-tekhnologij-v-borbe-s-pandemiej-covid-19>;
- Ефимов А. Р. и др. Экспериментальные исследования применимости беспилотных воздушных судов для решения актуальных задач корпоративной логистики // Полет. 2020. № 7. С. 15–22. URL: http://www.ros-polet.ru/files/archiv/pl720_web.pdf;
 - Ефимов А. Р. Актуальность использования промышленных экзоскелетов верхних конечностей для снижения количества профессиональных заболеваний опорно-двигательного аппарата // Медицина труда и промышленная экология. 2020. № 7. С. 412–416;
 - Semochkin A. N., Zabihifar S. and Efimov A. R. Object Grasping and Manipulating According to User-Defined Method Using Key-Points // 12th International Conference on Developments in eSystems Engineering (DeSE). 2019. Kazan, Russia, 2019. P. 454-459. DOI: 10.1109/DeSE.2019.00089. URL: <https://ieeexplore.ieee.org/document/9073432>;
 - Ефимов А. Р. Боевые роботы: угрозы учтённые или непредвиденные // Индекс безопасности. 2016. Т. 22. № 3-4. С. 79–96.

Глава 1. Методологический анализ современных исследований в интеллектуальной робототехнике

1.1. Обобщённый обзор современной интеллектуальной робототехники

Интеллектуальная робототехника представляет собой синтетическое направление современных исследований, в которых методы искусственного интеллекта применяются для симуляции роботами типов поведения свойственных интеллектуальному поведению человека.

Прежде всего, необходимо предложить рабочее определение Общего искусственного интеллекта. В недавнем техническом отчёте Европейского союза было рассмотрено 55 различных определений того, что составляет понятие «искусственный интеллект»²⁹. Обсуждение наилучшего из них не является целью настоящей работы, поэтому в качестве рабочего определения предлагается *определение Общего искусственного интеллекта*, разработанное на основе исследований Б. Герцля³⁰:

Общий искусственный интеллект — это способность обучаться и действовать лучше специалиста в любой области, достигая целей в широком диапазоне сред при ограничении ресурсов.

Проблема искусственного интеллекта в силу самой своей специфики оказывается на пересечении понятий физического и психического, затрагивает проблему телесности носителя интеллекта (включая те случаи, когда она приписывается машине). «Если мы принимаем в когнитивной науке т. н. телесно-воплощённый подход и тесно с ним связанный деятельностный подход (embodied and enacted approaches), то изменяется понимание целого ряда проблем»³¹. Сможет ли антропоморфный робот самостоятельно

²⁹ Samoili S., López Cobo M., Gómez E., De Prato G., Martínez-Plumed F. and Delipetrev B. AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. Luxembourg: Publications Office of the European Union, 2020. 90 p.

³⁰ Goertzel B. Artificial general intelligence: Concept, state of the art, and future prospects. // Journal of Artificial General Intelligence. 2014. Vol. 5. P. 1–46.

³¹ Лекторский В. А. Философия, искусственный интеллект, когнитивные исследования //

абстрагировать получаемый поток сигналов до уровня понятий? Изменится ли что-то качественно в процессе его познания, если управляющая им программа будет «в облаке», на удалённом сервере? И если абстрагирование будет успешным, то осуществится ли оно исключительно математическим анализом входящих данных или для познания мира машине необходим специфический носитель? По мнению В. А. Лекторского, если мы принимаем телесно-воплощённый подход, то нам «...придётся признать, что мир будет восприниматься по-разному в зависимости от того, как устроено познающее существо. Один и тот же реально существующий мир будет выступать как разный для пчелы, для блохи, для волка и для человека»³².

Обратимся к истории вопроса. Термин «робот» родился в 20-х годах прошлого века благодаря пьесе К. Чапека «РУР». Идея о том, что роботом может быть полезная для человека машина, сочетающая в себе механику и электронику, принадлежит А. Азимову. Его термин *robotics* (роботика, робототехника) стал общим названием семейства технологий и крупной научной области. Одновременно с адаптацией термина «робот» в академической среде развивалось направление «искусственный интеллект». Его становление началось на 100 лет раньше, чем возник сам термин. «Рассуждающие» машины создавались такими выдающимися изобретателями как Ч. Беббидж, С. Н. Корсаков, К. Цузе, А. Тьюринг и др. Первоначально элементная база — механическая — была ограничена, и только появление доступных электронных компонентов сделало возможным создание компьютеров, основанных на двоичной системе счисления и реализующих принципы алгоритмической вычислимости универсальной машины Тьюринга.

Рождением области исследований «искусственного интеллекта» принято считать фундаментальную статью А. Тьюринга «Вычислительные

Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (г. Москва, 17–18 марта 2016 г.) / Под ред. В. А. Лекторского, Д. И. Дубровского, А. Ю. Алексева. М., 2017. С. 42.

³² Там же. С. 42.

машины и разум», опубликованную в 1950 г. и поставившую эпистемологическую проблему «могут ли машины мыслить?». Однако термин «искусственный интеллект» был предложен Дж. Маккарти лишь в 1956 г.

Параллельно с исследованиями в области искусственного интеллекта получила развитие и робототехника. В силу крайней ограниченности элементной базы электроники и возможностей вычислительной техники, большинство задач механики и управления в робототехнике решались за счёт математических моделей механики и развития теории машин и механизмов. Лишь с появлением в конце 80-х годов быстрых микроконтроллеров, сенсоров и управляемых моторов, появилась возможность решения многих задач управления роботами путем вычислений, т. е. компьютеров. Это поставило перед исследователями междисциплинарную задачу по объединению возможностей искусственного интеллекта в решении задач робототехники и послужило началом появления «интеллектуальной робототехники» как отдельного направления исследований.

Из сказанного выше видно, что в самом начале современной истории искусственного интеллекта понятия робот и искусственный интеллект родились почти одновременно и были связаны общим понятийным комплексом — искусственный, механический человек. Однако из-за стремительного развития микроэлектроники и значительно более медленного развития материальной базы для создания роботов, эти научные области стали стремительно расходиться в стороны³³. Быстрый прогресс машин в работе с абстрактными понятиями в математике породил ожидания успеха в создании «сильного ИИ», подпитывал внимание к функционалистским концепциям. Постоянное увеличение мощности компьютеров, известное как эмпирический закон Мура, стало в определённой степени препятствием для осмысления телесности роботов подобно тому, как успехи физики И. Ньютона в XVIII-XIX вв. стали препятствием к философскому осмыслению онтологических

³³ Wooldridge M. The Road to Conscious Machines: The Story of AI. London: Penguin, 2020. P. 125–150.

противоречий в механистическом детерминизме. Термин «интеллектуальная робототехника» стал интегрирующим ответом, попыткой преодолеть междисциплинарные противоречия, растущие среди исследователей искусственного интеллекта и робототехники.

Впервые термин «интеллектуальная робототехника» начал появляться в научной литературе в начале 90-х годов прошлого века³⁴ (Г. Саридис, П. Дэвей). Первые упоминания «интеллектуальной робототехники» или «интеллектуальных роботов» связаны с идеей распределённого искусственного интеллекта: принятие решений можно реализовать непосредственно в исполнительном механизме — роботе, а не в «центральном компьютере», управляющим совокупностью роботов. Данный подход повторял ранее произошедшую в 70–80-х гг. революцию в широком распространении персональных компьютеров. Таким образом, интеллектуальная робототехника возникла как проникновение дисциплины «робототехника» в область исследований искусственного интеллекта, которая в конце 80-х годов прошлого века переживала период «зимы». Развитие искусственного интеллекта, основанного на логическо-символьном подходе, не принесло ожидаемых результатов, и смена парадигмы исследований была неизбежна. Кризисная ситуация с искусственным интеллектом отразилась в словах Р. Пенроуза, когда он написал в работе, вышедшей в 1989 г.: «хотя сделано было немало, произвести что-либо, достойное называться подлинным интеллектом, до сих пор никому не удалось»³⁵.

Многими исследователями предлагались синергетические междисциплинарные подходы для преодоления кризиса доверия к логико-символическим методам. Именно из-за необходимости дистанцироваться от провалов в исследования искусственного интеллекта и появилось семейство

³⁴ *Davey P.G.* Review: «3rd symposium on theory and practice of robots and manipulators», 1980 // *IEE PROC.* 1980. Vol. 128. Pt. D. No. 1. P. 18; *Saridis G.* Intelligent robotic control // *IEEE Transactions on Automatic Control.* 1983. Vol. 28. No. 5. P. 547–557.

³⁵ *Пенроуз Р.* Новый ум короля: О компьютерах, мышлении и законах физики / Пер с англ. В. О. Мальшенко. М.: УРСС, ЛКИ, 2011. С. 74.

терминов «интеллектуальные системы», «интеллектуальные роботы» и «интеллектуальная робототехника».

В новом направлении интеллектуальной робототехники сразу же возникло несколько новых интересных подходов к проблематике искусственного интеллекта. Поскольку полноценный процесс абстрагирования и конкретизации для компьютеров пока недостижим, то разработчики предлагали методы, основанные на использовании какого-то одного этапа абстрагирования, например, через описание всего поведения робота как набора чисто реактивных, рефлексивных действий. В последнее десятилетие прошлого века Р. Брукс³⁶ предложил применить бихевиористические методики к созданию искусственного интеллекта и роботов, которые он назвал «ситуативной робототехникой». По методологии Р. Брукса интеллектуальный агент помещается в конкретную физическую среду и действует в ней самостоятельно, опираясь лишь на заложенные в нём поведенческие правила, сенсорную систему и набор актуаторов. Это было полностью противоположно ранее принятому подходу, в котором экспертные системы, основываясь на логико-символическом подходе и репрезентациях окружающего мира, делали логические вычисления и принимали оптимальные решения, опираясь на запрограммированные знания. По мнению Р. Брукса³⁷, для построения любой системы искусственного интеллекта необходимо использовать робота, так как физическое воплощение для полноценного развития искусственного интеллекта обязательно по трём причинам: 1) интеллект проявляет себя в непосредственном взаимодействии с физическим миром; 2) интеллектуальное поведение, т. е. поведение, которое обладает сходными с человеком характеристиками, может воспроизводиться без использования эксплицитного знания об объекте, на который направлено действие интеллектуального агента; 3) интеллект может быть эмерджентным

³⁶ Brooks R. A. Intelligence Without Representation // Artificial Intelligence. 1991. No. 47. P. 139–159.

³⁷ Ibid. P. 139–159.

свойством агента, которое проявляется при взаимодействии агента и окружающей среды.

Р. Брукс и другие исследователи, подобно А. Тьюрингу, использовали операционалистский подход для того, чтобы избежать необходимости определений «что есть интеллект». По сути было предложено применение методов, созданных ранее в искусственном интеллекте, — сегментация и генерализация знаний для выбора наиболее подходящего ответа на вопрос «что есть робот?». Следуя этому подходу, любое техническое изделие человека, действующее в физическом мире, является интеллектуальным роботом, если обладает тремя способностями, выполняемыми одновременно:

1. Способностью получать данные об окружающем мире с помощью сенсорных приспособлений (видеокамеры и т. п.);

2. Способностью автономно (т. е. независимо от человека-оператора) строить модели своего поведения, выбирая оптимальные пути решения поставленной задачи на основе динамически адаптируемых моделей поведения;

3. Способностью производить действия в физическом мире путём манипуляции объектами физического мира и/или собственного перемещения.

Следуя Р. Мёрфи³⁸, можно суммировать определение интеллектуального робота, как интеллектуального агента, действующего в физическом мире, а три упомянутых свойства — как его характеристики.

Рассмотрим более детально составляющие верхнеуровневые элементы интеллектуального робота как интеллектуального агента. В общем случае такой робот состоит из пяти функций:

1. Сенсоры (sense) — функции, которые обеспечивают восприятие роботом окружающего мира, подобно органам чувств человека. Эти устройства обладают способностью преобразования физических измерений к

³⁸ *Murphy R. Introduction to AI Robotics. Cambridge: The MIT Press, 2019. P. 56.*

семантически значимой информации. Например, лазерные дальнометры указывают роботу расстояние до препятствия.

2. Понимание (think) — эта функция обеспечивает обработку сенсорной информации, построение моделей окружающего мира, адаптацию этих моделей в зависимости от окружения и принятие решения относительно тех или иных действий в окружающем робота физическом мире. Например, робот имеет в памяти цифровую карту-модель пространства, в котором он должен передвигаться, уклоняясь от динамических препятствий (людей).

3. Действие (act) — эта функция обеспечивает роботу возможность изменения физического мира вокруг. Действия робота в окружающей среде основаны на выполнении элементарных действий актуаторами (манипуляторы, колеса).

4. Коммуникация (communications) — робот должен обладать функцией коммуникации с человеком или другими роботами.

5. Энергообеспечение (energy) — робот обладает возможностью для выполнения функций 1–4 будучи полностью энергонезависимым в период выполнения задачи.

Завершая краткий обобщённый обзор интеллектуальной робототехники, необходимо осветить вопросы актуальности её применения в жизни человека, а также перспективы дальнейшего развития. Современное общество является постиндустриальным в силу того, что высока доля сервисного сектора в общем объёме производства³⁹. В предыдущие десятилетия рост робототехники обеспечивался прежде всего за счёт того, что роботов для промышленности делать проще, — можно относительно легко адаптировать промышленное окружение для робота, а не робота под промышленное окружение. Однако промышленность, как точка роста, имеет своё ограничение. Поэтому и дальнейший рост робототехники будет происходить за счёт сервисной

³⁹ CIA Fact Book. URL: <https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html> (дата обращения: 02.06.2020).

робототехники, в которой находят массовое применение интеллектуальные роботы⁴⁰.

На текущий момент интеллектуальные роботы используются для удалённого присутствия и развлечений. Многие авторы начинают рассмотрение применения робототехники с деятельности трёх видов, которая в экономической литературе получила название 3D: dirty (грязная), dull (скучная), dangerous (опасная)⁴¹. Многочисленные примеры использования интеллектуальных роботов приведены в ежегодных отчётах Лаборатории робототехники Сбербанка. Роботизированные хирургические комплексы Intuitive Surgical имеют разрешение на проведение эндоскопических операций нескольких типов. По данным на конец 2018 г. почти 5000 роботов-хирургов по всему миру провели более миллиона операций разного типа⁴². Одним из важнейших видов интеллектуальных роботов являются автономные транспортные средства AGV/AMR (Automated Guided Vehicles/Autonomous Mobile Robots): в 4–5 раз сокращают по времени операции перемещения грузов⁴³. В качестве примера интеллектуального робота для логистических операций можно привести робота-курьера Jeeves. Данная платформа обладает способностью автономно прокладывать маршрут перемещения в многоэтажных зданиях, пользуясь при этом лифтами, так же она оснащена трёхмерным зрением и набором сенсоров.

⁴⁰ Лаборатория робототехники Сбербанка. Аналитический обзор мирового рынка робототехники. URL: https://www.sberbank.ru/common/img/uploaded/pdf/sberbank_robotics_review_2019_17.07.2019_m.pdf (дата обращения: 21.07.2020).

⁴¹ *Connell J. Kitanai, Kitsui and Kiken: The Rise of Labour Migration to Japan // Economic & Regional Restructuring Research Unit. Sydney: University of Sydney, 1993. P. 14.*

⁴² Лаборатория робототехники Сбербанка. Аналитический обзор мирового рынка робототехники. URL: https://www.sberbank.ru/common/img/uploaded/pdf/sberbank_robotics_review_2019_17.07.2019_m.pdf (дата обращения: 21.07.2020).

⁴³ Там же.

Рынок интеллектуальных роботов для персональных нужд (edutainment) растёт на 10–30% в год⁴⁴. Люди постоянно нуждаются в развлечениях, обучении, уборке помещений или услугах безопасности.

Таким образом, можно предположить, что пределы роста распространения интеллектуальных роботов в современной цивилизации ограничены лишь пределами сервисной экономики в глобальном масштабе. Это открывает невероятно широкие перспективы для применения интеллектуальных роботов в сочетании с Общим искусственным интеллектом.

1.2. Основные тренды современной интеллектуальной робототехники в свете методологических проблем создания Общего искусственного интеллекта

Робототехника и искусственный интеллект являются разделами техники и поэтому для них значимы определения и тренды осмысления понятия технического в целом. Проводя инвентаризацию фундаментальных и классических понятий техники на основе идей Э. Каппа, О. Шпенглера, Л. Мэмфорда, М. Хайдеггера, С. Шавиро и С. Бескаравайного мы выделим следующие аспекты, повлиявшие на развитие искусственного интеллекта и робототехники.

Во-первых, техника, в том числе робототехника, включая искусственный интеллект, раньше не воспринималась философами как приоритетное направление. Во второй половине прошлого века, несмотря на бурное развитие робототехники в отношении космических, атомных технологий или промышленности, робототехника всегда уделяла основное внимание классическим естественно-научным проблемам.

Во-вторых, развитие философии искусственного интеллекта в русле функционалистской парадигмы было во многом направлено на исследование природы сознания человека. Философы, такие как Х. Патнэм, Дж. Сёрл, Т. Нагель, Д. Деннет (см. раздел 1.3), использовали искусственный интеллект

⁴⁴ Там же.

(или роботов) как мысленный инструмент для построения конструкторов, проясняющих теорию сознания, решающих «трудную проблему сознания». Философия искусственного интеллекта и робототехники оставалась во многом за скобками философского мейнстрима, несмотря на значительное количество работ, посвящённых этой теме.

В-третьих, в общей научной парадигме техника воспринималась как производная от науки. При этом наука постоянно использует новые технические инструменты для получения новых знаний. До последнего времени научное сообщество не уделяло достаточного внимания тому факту, что искусственный интеллект может быть самостоятельным фактором в научных открытиях. Сегодня же руководитель программы по созданию искусственного интеллекта корпорации Sony Х. Китано считает, что тест Тьюринга больше не может быть критерием создания искусственного интеллекта, а на существующем уровне техники возможно разработать «систему искусственного интеллекта, которая может самостоятельно сделать открытия в области биомедицины, которые будут настолько прорывными, что будут удостоены Нобелевской премии»⁴⁵.

В первой половине XXI века разрыв между образом фундаментальной науки в глазах общества и стратегическими приоритетами развития технологий для нужд государства и общества стал еще больше. Техника, делающая упор на искусственный интеллект и робототехнику, становится инструментом геополитики, постепенно вытесняя поисковую парадигму исследований прикладными задачами. Одним из подтверждений этого тезиса служит внимание государства к области искусственного интеллекта. Согласно заявлению Президента России В. В. Путина, «Как и великие географические открытия, как выход в космос, это ещё и шаг в неизведанное. Особенно это касается развития так называемого сильного искусственного интеллекта, который может стать фундаментом и прорывом для всей цивилизации». По

⁴⁵ *Kitano H. Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery // AI Magazine SPRING. 2016. P. 41.*

сути, этим высказыванием глава российского государства задал тематике искусственного интеллекта и робототехнике стратегический характер, отмечая, что «Потенциал применения технологий искусственного интеллекта огромен, причём абсолютно в любых направлениях и отраслях»⁴⁶. Подобные высказывания можно найти и у лидеров других государств. Это свидетельствует о том, что робототехника и искусственный интеллект не просто являются цивилизационными приоритетами, но сама цивилизация уже не мыслится без развития этих областей.

Основателем философии техники принято считать Э. Каппа, который предложил концепцию «органопроекции» и утверждал, что «как продукт деятельности мозга и руки, орудие находится в таком глубоком внутреннем средстве с самим человеком, что он в создании своей руки видит объективированным перед своими глазами нечто от своего собственного «я»... часть самого себя». Хотя концепция неоднократно критиковалась, она во многом предвосхитила такие направления робототехники как антропоморфные роботы, а также роботы телеприсутствия, которые становятся продолжением-расширением человека. Именно такова концепция робота-аватара, который создается компанией Toyota,⁴⁷ — робот, обладающий подвижностью человеческого тела и способный к обмену с оператором тактильной информацией.

Понимание любых задач, стоящих перед создателями искусственного интеллекта, невозможно без хотя бы приблизительного ответа на вопрос о том, что есть «естественный интеллект». Исследование естественного интеллекта человека является само по себе важным научным направлением нейро- и когнитивных наук. По мнению пионера отечественной школы изучения проблем интеллекта человека М. А. Холодной, существует три

⁴⁶ Конференция по искусственному интеллекту. URL: <http://kremlin.ru/events/president/news/62003> (дата обращения: 20.07.2020).

⁴⁷ Why is Toyota Developing Humanoid Robots? URL: <https://global.toyota/en/newsroom/corporate/30609642.html> (дата обращения: 26.07.2020).

принципиального подхода к осмыслению естественного интеллекта в контексте создания искусственного интеллекта. В рамках первого подхода, «...постулируется необязательность использования психологических знаний об особенностях устройства естественного интеллекта при конструировании системы искусственного интеллекта»⁴⁸. Это сходно с известным утверждением о том, что полёт птицы и полёт самолета основываются на различных физических законах, и выводом о том, что искусственный интеллект не должен быть основан на известных нам принципах работы естественного интеллекта. Косвенно это подтверждается недавней победой компьютерной программы AlphaGo над сильнейшим в мире игроком в го. По заявлениям наблюдателей, ряд ходов компьютера были «не человеческими»⁴⁹. Второй подход основывается на том, что «предметом моделирования в искусственном интеллекте выступают некоторые частные свойства человеческого интеллекта»⁵⁰. Этот подход является одной из составляющих функционализма. Третий подход, следуя Р. Пенроузу и Дж. Сёрлу, утверждает «невозможность создания искусственного интеллекта как аналога естественного интеллекта в силу их принципиальной несопоставимости»⁵¹. Компьютерная программа, согласно Дж. Сёрлу, не может обладать ментальностью, а значит и не может служить аналогом интеллекта в деле создания его искусственного аналога.

В настоящее время именно второй подход, согласно которому некоторые способности естественного интеллекта могут служить аналогами некоторых свойств искусственного, является главенствующей парадигмой

⁴⁸ *Холодная М. А.* Естественный и искусственный интеллект. Структура и функции естественного интеллекта в контексте проблемы искусственного интеллекта // Искусственный интеллект. Междисциплинарный подход / Под ред. Д. И. Дубровского, В. А. Лекторского. М.: ИИнтелл. 2006. С. 149.

⁴⁹ How Google's AI Viewed the Move No Human Could Understand. URL: <https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/> (дата обращения: 29.07.2020).

⁵⁰ Там же. С. 149.

⁵¹ Там же.

создания систем искусственного интеллекта и интеллектуальной робототехники. Среди способностей естественного интеллекта В. К. Финн выделяет следующие тринадцать ключевых способностей, которые важны для разработки искусственного аналога:

1. «Способность выделять существенное в наличных знаниях, т. е. упорядочивать их (необходимый аспект интуиции).

2. Способность к целеполаганию и планированию поведения — порождение последовательности “цель → план → действие”.

3. Способность к отбору знаний (посылок, выводов, релевантных цели рассуждения).

4. Способность извлекать следствия из имеющихся знаний, т. е. способность к рассуждению, которое может содержать как правдоподобные выводы, используемые для выдвижения гипотез, так и достоверные выводы.

5. Способность к аргументированному принятию решений, использующему упорядоченные знания (представление знаний) и результаты рассуждений, соответствующие заявленной цели.

6. Способность к рефлексии — оценке знаний и действий.

7. Наличие познавательного любопытства: познающий субъект должен быть способен задавать вопрос “что такое?” и искать на него ответ.

8. Способность и потребность находить объяснение (не обязательно продуктивное) как ответ на вопрос “почему?”.

9. Способность к синтезу познавательных процедур, образующих эвристику решения задач и рассмотрения проблем, например, такой является взаимодействие индукции, аналогии и абдукции (с учётом фальсификации выдвигаемых гипотез посредством поиска контрпримеров) с последующим применением дедукции.

10. Способность к обучению и использованию памяти.

11. Способность к рационализации идей, стремление уточнить их как понятия.

12. Способность к созданию целостной картины относительно предмета мышления, объединяющей знания, релевантные поставленной цели.

13. Способность к адаптации в условиях изменения жизненных ситуаций и знаний, что означает коррекцию «теорий» и поведения»⁵².

В. К. Финн считает, что способности 1, 3, 4, 5, 6, 8, 9 и 10 — могут быть воплощены в полном объёме достаточно развитой вычислительной системой (к примеру, на базе универсальной машины Тьюринга). Для их осуществления не нужны будут операторы, т.е. эти способности могут быть реализованы в полностью автономном режиме. Но способности 2, 7, 12 и 13 недостижимы для машины на уровне, эквивалентном человеческому, без участия человека-оператора, т. е. их реализация возможна только лишь когда человек становится супервизором, действуя вместе с машиной в интерактивном режиме. Говоря словами Х. Дрейфуса, «то, чего никогда не сможет выполнить компьютер» на уровне человека — это задавать цели, задавать вопрос «что это такое?», а в поисках ответа — создавать целостную картину мира и полностью корректировать поведение в зависимости от условий внешнего мира.

Вопрос о том, что может сделать робот (искусственный интеллект) и что ему сделать невозможно (что навсегда останется ответственностью человека) перекликается с вопросом о том, в чём состоит разница между техническими устройствами, которые составляют искусственный интеллект (робототехнику) и биологическими системами. Мы разберём эту разницу в том плане, который С. Бескаравайный называет «парадигмой техники»,⁵³ и поясним далее, в каких из этих направлений возможны резкие скачки. Среди основных трендов, влияющих на современную интеллектуальную робототехнику и её вклад в развитие Общего искусственного интеллекта, необходимо отметить следующие:

⁵² Финн В. К. Искусственный интеллект: Методология, применения, философия. М.: Красанд, 2018. С. 37–39.

⁵³ Бескаравайный С. С. Бытие техники и сингулярность. М.: РИПОЛ Классик, 2018. С. 209.

Конструкционный материал и химическая основа. Самые сложные работы на текущий момент имеют около 50 степеней свободы. Однако человек имеет около 244 степеней свободы. В млекопитающих присутствует лишь шесть основных химических элементов и около двадцати элементов в остаточном количестве. Для робота число химических элементов, которые используются для изготовления устройств, — свыше 50. Разнообразием синтетических материалов инженеры вынуждены замещать биохимические механизмы работы клетки. Использование *нитрида галлия, мягких кристаллов, графена* — открывает новые возможности по имитации организма.

Новые источники энергии, технологии хранения и сбора электроэнергии. Нынешнее поколение самых совершенных роботов имеет в своем распоряжении разнообразные двигатели. По доступной энергии они намного превышают возможности человека, но эффективность метаболизма биологических систем намного превышает роботов и компьютеры. Для создания распространённых антропоморфных машин необходимо кардинально улучшить их энергетические характеристики. Это можно сделать за счёт *развития существующих технологий хранения энергии или исследований в области альтернативных источников энергии (например, солнечные панели для беспилотных летательных аппаратов, портативные водородные элементы питания).*

Возможности обработки информации. В основе всей компьютерной отрасли лежит бинарный код, основанный на булевой алгебре. При этом доминирующая архитектурная парадигма компьютеров — «фон неймановская архитектура» — основана на постоянном разделении памяти и вычисления. Биологические системы имеют принципиально другое устройство, в том числе из-за других принципов вычислений. Прежде всего, это асинхронный механизм обмена информацией между нейронами с помощью синапсов. Так

же мозг имеет высочайшую плотность вычислительных узлов по сравнению с компьютером⁵⁴.

Машинное обучение для роботов. Робототехника последних десятилетий, несомненно, развивалась под воздействием статистических методов и машинного обучения с упором на искусственные нейронные сети глубокого обучения. По мнению главного исследователя Лаборатории робототехники компании Google В. Кумара, «машинное обучение помогает нам преодолевать технологические вызовы практического использования дешёвых манипуляторов»⁵⁵. В частности, исследователи Лаборатории робототехники Сбербанка видят следующие пути повышения эффективности искусственных нейронных сетей: 1) сохраняя приемлемую скорость обучения искусственной нейронной сети, необходимо усложнять её архитектуру и увеличивать ёмкости; 2) снижать энергопотребление систем «машина — нейронная сеть» при режиме работы onboard⁵⁶.

Человеко-машинное взаимодействие. *Робот как инструмент, повторяющий возможности человека,* — экзоскелеты, экзопротезы и нейрокомпьютерные интерфейсы, которые в результате могут привести к появлению нового вида человека Homo Extensis (человек дополненный).

Робот как инструмент, расширяющий возможности человека. Именно об этом говорили ещё древнегреческие мифы: помощниками бога-кузнеца Гефеста были специальные умные треножники, которые вовремя подавали ему нужные инструменты.

Аватаризация является важнейшим направлением робототехники. Освоение космоса или глубин мирового океана невозможно без создания

⁵⁴ Koch C. The Feeling of Life Itself. Cambridge: MIT Press, 2019. P. 35–73.

⁵⁵ Inside Google's Rebooted Robotics Program, The New York Times, 2019. URL: <https://www.nytimes.com/2019/03/26/technology/google-robotics-lab.html?smid=nytcore-ios-share> (дата обращения: 28.07.2020).

⁵⁶ Лаборатория робототехники Сбербанка. Аналитический обзор мирового рынка робототехники. URL: https://www.sberbank.ru/common/img/uploaded/pdf/sberbank_robotics_review_2019_17.07.2019_m.pdf (дата обращения: 21.07.2020).

роботов-аватаров, которые могут дистанционно управляться человеком и синхронно повторять его движения.

Специфика повседневной жизни и профессиональной деятельности, которая возникает при взаимодействии роботов и людей, становится предметом разнообразных исследований. Голосовые помощники и чат-боты — это пример успешной реализации технологий социального взаимодействия.

Взаимодействие групп роботов и людей. Известны различные виды группового поведения роботов — стайное, роевое. В таких группах система управления создаёт единое информационное поле для управления группой дронов. Важно отметить, что в процессе эксплуатации систем группового управления автономность отдельного дрона повысится: каждый участник группы сможет расширять собственные возможности связи и информационного обмена за счёт других участников.

Манипуляции объектами в физическом мире. Для того чтобы эффективно действовать в мире природы или искусственных предметов, роботы должны обладать манипуляционными способностями, которые приближаются к способностям животных или людей. Человек создаёт величайшее разнообразие форм окружающих его предметов, каждый из которых имеет определённые атрибуты, которые оцениваются людьми в доли секунды для обеспечения возможности манипуляции ими. Роботы пока не обладают такими способностями. Как подчёркивается в национальной стратегии США по робототехнике, «гибкие механизмы захвата, которые смогут позволить манипуляции с повседневными предметами, являются национальным приоритетом»⁵⁷.

Главным направлением исследований в области манипуляции является использование методов машинного обучения для обработки информации, которая поступает от сенсоров давления. Кроме того, новые возможности

⁵⁷ Robotics Roadmap and the National Robotics Initiative 2.0 (2016). URL: <https://cra.org/crn/2017/02/2016-robotics-roadmap-national-robotics-initiative-2-0/> (дата обращения: 12.07.2020).

роботов для манипуляции предметами приходят из исследований в области «мягкой робототехники»⁵⁸.

Сенсорика и органы восприятия окружающей среды. *Развитие модульных решений*, которые включают набор аппаратного и программного обеспечения для разработки роботов разного уровня сложности. Такие решения включают инструменты для обеспечения распознавания объектов, визуальной локализации и т. д. *Улучшение сенсорных возможностей роботов по ряду метрик*: точность распознавания мимики, жестов, эмоций, шумов.

В заключении данного параграфа необходимо упомянуть о важных исследованиях в области нейробиологии и сознания, включая экспериментальные исследования А. Дамасио, Дж. Риццоллати, К. Коха, В. Рамачандрана и др. В частности, А. Дамасио подчеркивает, что вопросы вычислимости не являются определяющими при создании искусственного интеллекта, способного к познанию мира на уровне, сопоставимом с человеком. По его мнению, способности, связанные с восприятием, движением, памятью, владением языком могут быть вычислимы и являются так называемыми способностями первой производной, но общие познавательные способности связаны с эмоциями, которые намного шире, чем вычисляемые способности и поэтому являются способностями второй производной⁵⁹. Это хорошо коррелирует с тем, что ранее указывалось в связи с определением естественного интеллекта по В. К. Финну. Усиливая вышесказанное, А. Дамасио утверждает в своей книге, что самость является частным случаем интеллекта, которая ощущается и которая картируется в каждом возможном сенсорном опыте — визуальном, аудиальном,

⁵⁸ *Billard A., Kragic D.* Trends and challenges in robot manipulation // Science 364 (6446), eaat8414. URL: <https://science.sciencemag.org/content/364/6446/eaat8414> (дата обращения: 12.07.2020).

⁵⁹ *Gardels N.* USC neuroscientist Antonio Damasio argues... URL: <https://ngp.usc.edu/usc-neuroscientist-antonio-damasio-argues-that-feelings-and-emotions-are-what-make-up-human-intelligence-consciousness-and-the-capacity-for-cultural-creation-a-map-of-the-computational-mind-he-says/> (дата обращения: 12.07.2020).

осязательном, мышечном и висцеральном. Другими словами, сознание — результат комплексной реакции на стимулирование телесных функций⁶⁰.

Другим важнейшим результатом оказались зеркальные нейроны, открытые нейробиологом Дж. Риццолатти, которые могут дать новые возможности для имитационного обучения нейронных сетей и роботов⁶¹, так как являются основой для сенсорно-моторных активностей приматов и человека. В соответствии с ранее принятым допущением о том, что отдельные функции искусственного интеллекта могут быть моделированы на основе архитектуры естественного интеллекта, зеркальные нейроны могут быть неплохим кандидатом на эту роль, так как имеют значимость не только для моторики, но и языковых способностей человека⁶².

Одной из лидирующих теорий сознания в нейронауке настоящего времени является Интегрированная Информационная Теория (ИТ), авторами которой являются нейробиологи Дж. Тонони и К. Кох. В соответствии с этой теорией, сознание может возникнуть в любой достаточно сложной самоорганизующейся системе. Компьютеры и роботы сегодняшнего уровня развития техники имеют достаточно низкий уровень сложности в сравнении с такого рода системами.

Нейропсихолог и невролог В. Рамачандран высказал гипотезу о процессе «межмодальной абстракции»⁶³, которой он называет «способность вычислять сходство» в различных типах восприятия. Эта способность, «несмотря на внешние различия, могла проложить путь к более сложным типам абстракции, которыми успешно пользуется наш человеческий вид». По

⁶⁰ Дамасио А. Я, мозг и возникновение сознания / Пер. с англ. И. Ющенко. М.: Карьера Пресс, 2018. 384 с.

⁶¹ *Maistros G., Hayes G. Towards an Imitation System for Learning Robots // Vouros G.A., Panayiotopoulos T. (eds.) Methods and Applications of Artificial Intelligence: SETN 2004. Lecture Notes in Computer Science. Vol. 3025. Berlin, Heidelberg: Springer, 2004. URL: https://doi.org/10.1007/978-3-540-24674-9_26 (дата обращения: 23.07.2020).*

⁶² Черниговская Т.В. Зеркальный мозг, концепты и язык: цена антропогенеза // Российский физиологический журнал. 2006. № 1. С. 84–99.

⁶³ Рамачандран В. Мозг рассказывает. М.: Карьера Пресс, 2005. С. 124.

мнению Рамачандрана, именно «зеркальные нейроны, возможно, оказались эволюционным проводником, благодаря которому это произошло»⁶⁴.

Расшифровка мозговых кодов может оказать кардинальное влияние не только на исследования в области нейронаук и нейрофизиологии человеческого мозга, но и на исследования в области искусственного интеллекта в силу описанной выше связи между эмоциональным и сенсорно-моторным контурами сознания. Исследования по этой тематике начал развивать в СССР философ Д. И. Дубровский, изложивший основы подхода к расшифровке мозговых кодов психических явлений⁶⁵ задолго до того, как эта проблема стала популярной за рубежом. В этой книге были впервые предложены практические шаги по картированию нейродинамических кодов и методологии их расшифровки (эта книга, вышедшая 50 лет тому назад, недавно переиздана⁶⁶). Современные исследования в этом направлении имеют фундаментальное значение для нейронаук и теории Общего искусственного интеллекта.

Можно сказать, что приближение к образу мыслящего робота идёт эволюционно, — инженерам и мыслителям проще сказать, чем не должен быть робот и гораздо труднее охарактеризовать его сходство и различия с человеком.

1.3. Теоретические и эпистемологические трудности развития интеллектуальной робототехники и Общего искусственного интеллекта

Интеллектуальная робототехника и Общий искусственный интеллект являются частью грандиозной по своим масштабам технической революции, которая, по словам В. С. Степина «основана на переходе к сложным системам,

⁶⁴ Там же. С. 124.

⁶⁵ *Дубровский Д. И.* Психические явления и мозг: философский анализ проблем в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М.: Наука, 1971. 386 с.

⁶⁶ *Дубровский Д. И.* Проблема сознания. Теория и критика альтернативных концепций. М.: ЛЕНАНД, 2020. 400 с.

в том числе системам саморегуляции и саморазвития»⁶⁷. На сегодняшний день ни робототехника, ни искусственный интеллект не приблизились к тому, чтобы создать универсальную машину, способную к полностью автономной работе в достаточной широкой области, которую можно было бы сравнить с деятельностью человека. Нерешённые практические трудности всё более обнажают глубочайшие теоретические и эпистемологические проблемы, связанные с созданием интеллектуальных роботов и Общего искусственного интеллекта. Робот или в общем случае вычислительное устройство, снабжённое актуаторами, имеет объективные характеристики (программа, физика его манипуляторов или двигательной платформы), однако восприятие его человеком определяется деятельностными возможностями самого робота. Отсюда вытекает проблема субъектности робота или носителей искусственного интеллекта.

Но базовой проблемой становится проблема целостности, так как робот фрагментирован: а) отсутствует неразрывная связь физического и психического, характерная для человека и его деятельности, есть возможность удаленного управления роботом, программой на сервере; б) он может проводить лишь часть процесса абстрагирования-конкретизации (инструменты, которые эволюционно сформировались в организме и присутствуют в культуре, у робота присутствуют лишь для выполнения определённых рабочих функций); в) не обладает целостной картиной мира; г) ограничен в накоплении опыта — перезагрузка может уничтожить накопленные данные.

Главная линия демаркации современной теории и практики исследований в области искусственного интеллекта, интеллектуальной робототехники заключается в роли чувственного восприятия и роли

⁶⁷ *Степин В. С.* Постнеклассическая рациональность и информационное общество // *Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (17–18 марта 2016 г., г. Москва, философский факультет МГУ им. М.В. Ломоносова).* М.: ИИнтелл, 2017. С. 59–70.

репрезентаций и генерализаций знаний для действий роботов (или искусственного интеллекта в общем смысле) в физическом мире. Обсуждение этой проблематики началось еще с работы британского математика А. Тьюринга, который отказался рассматривать «воплощённый интеллект» в «электрическом теле», но однако признавал превосходство «телесности» над логико-символическим подходом для обучения робота. В работе «Intelligent Machinery» 1948 г. он впервые проанализировал ограничения и возможности «воплощённого искусственного интеллекта» при попытке имитировать человека. Характерным ограничением, на которое указывал А. Тьюринг, было то, что, построив «создание», имитирующее человека, конструкторы столкнутся с тем, что «отсутствие доступа этого создания к простым человеческим радостям, таким как еда, секс, спорт и многие другие, создаст препятствия для имитации интеллекта»⁶⁸, но сама форма этого утверждения скорее визионерская, чем аналитическая.

При всём богатстве научной терминологии философия искусственного интеллекта, как отдельная дисциплина научного мышления, сформировалась на десять-пятнадцать лет раньше, чем вошел в оборот термин «искусственный интеллект». Становление первой философской концепции искусственного интеллекта принято связывать с именем А. Тьюринга⁶⁹. В начальный период развития цифровых вычислителей участники Ratio Club, который возглавлял А. Тьюринг, считали, что для решения проблемы думающих машин требуется не создание какой-то специфической дисциплины, но опора на весь спектр научных дисциплин – от естественно-научных (химия, физика, математика) до гуманитарных (философия, психология, педагогика).

Это старая проблема соотношения универсальности и специализированности научного знания, которая обостряется при создании

⁶⁸ Turing A. M. Intelligent Machinery // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 508.

⁶⁹ Turing A. Computing machinery and intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 562–569.

любой дисциплины. Когда границы нового направления исследований лишь очерчиваются, нет строгого категориального аппарата, нет списка куновских «стандартных головоломок» и, кажется, что лишь энциклопедическое образование может помочь специалисту.

Говоря метафорически, А. Тьюринг создал философский субстрат, среду, отчасти дискурс, но понятие «искусственный интеллект» выделили из неё уже другие исследователи. Дж. Маккарти, математик-исследователь, предложил в 1956 г. провести двухмесячный семинар по теме «искусственного интеллекта» в Дартмутском колледже. Сейчас отчасти наивными кажутся стремления его коллег решить за это время большинство вопросов создания искусственного интеллекта, включая переводы текстов с различных языков, решение абстрактных математических задач, ремонта компьютеров, распознавания образов на уровне человека⁷⁰, — фактически уже рассматривая эти проблемы как «головоломки» в куновском смысле.

При этом сам термин «искусственный интеллект» является буквально «словом-чемоданом», куда каждый исследователь готов вложить собственный смысл. По данным сайта philpapers.org насчитывается 53 подкатегории общего раздела «Философия искусственного интеллекта». Хотя очень многие работы, затрагивающие эти темы, находятся в разделе «Философия сознания».

Практически все современные эпистемологические концепции искусственного интеллекта основаны приблизительно на пяти различных философских воззрениях, которые эволюционировали во второй половине XX века и были, помимо А. Тьюринга (1950), развиты представителями аналитической философии: Х. Патнэм (1960), Т. Нагель (1974) и Дж. Сёрл (1980), Д. Деннет (1997) и Дж. Чалмерс (1996).

По определению, приведённому в Кембриджском словаре, искусственный интеллект — это изучение методов производства машин,

⁷⁰ *McCarthy J.* A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. URL: <http://www-normal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (дата обращения: 26.04.2020).

которые имеют ряд качеств, таких как ум человека, способность понимать язык, распознавать образы, решать проблемы и обучаться⁷¹. В научноцентрической европейской философии, начиная с Нового времени, равно как в науке и технике, создание машин, способных действовать и рассуждать подобно человеку, рассматривалось как желаемое и мыслимое. Создание таких машин, по мнению авторов, первых автоматов XVI века или механических вычислителей XIX века требовало лишь корректно работающих механических частей, которые будут имитировать действия человека. Но только А. Тьюринг соотнёс новые свойства элементарных цифровых машин и мышление человека.

А. Тьюринг в 1936 г. сумел подставить в теорему Геделя о неполноте вместо арифметического языка простые воображаемые машины и доказал, что с помощью полученной «машины Тьюринга» можно вычислить всё, что только может быть вычислено по определённому алгоритму.

А. Тьюринг был неформальным лидером группы ученых, которая называла себя Клуб Рацио (*Ratio Club*). Эта группа встречалась примерно раз в месяц с сентября 1949 г. по июнь 1954 г. Своим названием клуб был обязан идее Н. Винера, который в одной из своих книг упомянул идею Лейбница о вычисляющей машине *Machina Ratiocinator*⁷². Среди обсуждаемых тем были: «Обучение цифрового компьютера» и «Проблемы роста», «Симуляция обучения», «Телепатия», «Дискриминационная логика» и др.⁷³

Два подхода концептуализации мыслящих машин, которые стали использоваться в исследованиях искусственного интеллекта, ведут своё начало из идей А. Тьюринга⁷⁴. Во-первых, какие методы, идеи, подходы

⁷¹ Cambridge English Dictionary // Artificial Intelligence. URL: <https://dictionary.cambridge.org/ru/словарь/английский/artificial-intelligence> (дата обращения: 05.05.2020).

⁷² *Warwick K., Shah H.* Turing's Imitation Game: Conversations with The Unknow. Cambridge: Cambridge University Press, 2016. 202 p.

⁷³ Там же.

⁷⁴ *Turing A.* Computing machinery and intelligence // *Alan Turing: His Work and Impact* / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 562–569.

можно использовать, чтобы создать думающую машину? Как можно узнать, что эта машина действительно думает не хуже, а, возможно, и лучше человека? Во-вторых, как можно использовать подход, основанный на универсализме машины Тьюринга, в изучении мышления и сознания в онтологическом и эпистемологическом значении?

Важнейшая составляющая гипотез А. Тьюринга об искусственном интеллекте — это отказ от дескриптивного подхода к мыслящим машинам. Если невозможно в исторически обозримом периоде решить проблемы мышления и сознания, то надо просто убрать «машину» из описания. Требуется мысленный эксперимент, суть которого настолько проста, что интуитивно понятна наблюдателю, причем мерой мышления в этом эксперименте выступает опять-таки человек: «если человек не может отличить интеллектуальное поведение машины и человека, то можно ли сказать, что машина наделена мышлением и интеллектом?». Игра в имитацию, по мнению А. Тьюринга, стала лучшей формой проведения такого эксперимента. Если человек-наблюдатель не может различить машину и человека, опираясь лишь на их деятельность, но не принимая во внимание внешний вид, то компьютер (думающая машина) может мыслить. Сформулированные в предыдущие десятилетия понятия универсальной машины и вычислимости алгоритмов А. Тьюринг сумел применить к философским проблемам сознания и тела (материи). Прямая связь математики и мышления машины, вероятно, составляет наиболее сильную часть бихевиористского подхода А. Тьюринга к сумме проблем, которые позднее станут наукой об искусственном интеллекте. Связь формального и казуального обеспечивается машиной Тьюринга и создаёт базу для концепции функционализма в философии сознания. Дихотомии функциональное/реальное, когниция/реализация, формальное/казуальное были вскрыты благодаря развитию функционалистских гипотез А. Тьюринга⁷⁵.

⁷⁵ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИнтеЛЛ, 2013. С. 204.

Критиковать концепцию «мыслящих машин» А. Тьюринга начали уже его современники. К числу наиболее радикальных критиков можно отнести К. Поппера, который называл компьютеры «карандашами» и потому даже не имел отдельных работ, посвященных им. Он писал: «при всём глубочайшем уважении к Тьюрингу, я не могу согласиться с ним, что компьютеры могут мыслить... то обстоятельство, что компьютер может делать больше из того, что делаем мы, обязано факту, что именно мы вкладываем в компьютер мощные операционные силы, в сущности, принципы автономного мира три»⁷⁶.

Если А. Тьюринг поставил вопрос о мышлении машин, то американский философ Х. Патнэм переосмыслил идеи А. Тьюринга в приложении к философии сознания, с позиций функционального подхода и использования формализма машин Тьюринга.

Основания философского направления, которое позднее стали именовать машинным (компьютерным) функционализмом, сформулировал именно Х. Патнэм. Он рассматривал любую сознательную деятельность как совокупность работы машин Тьюринга, выполняющих свои инструкции, — все это иногда именуют компьютерной метафорой. Он проводил прямую аналогию между мозгом и разумом в сравнении с аппаратным обеспечением компьютеров и их программами (статья «Сознание и машины»⁷⁷): интеллект — это своего рода программа, она не может быть материальной.

Х. Патнэм в рамках функционализма развивает идею о том, что важен вид, функциональная связность совершаемых операций, а не физическая природа машины, которая их совершает. Биологическую машину можно заменить технической (физической). Если функция в обоих процессах остается неизменной, результаты неразличимы, то их можно считать тождественными согласно принципу тождества. Функция, по Х. Патнэму,

⁷⁶ Popper K., Eccles J. *The Self and Its Brain: An Argument for Interactionism*. Berlin: Springer, 1977. P. 208.

⁷⁷ Putnam H. *Minds and Machines // Dimensions of Mind* / Ed. by S. Hook. N.Y.: New York University Press, 1960. P. 148–179.

может иметь множество реализаций. Одно и то же действие может быть воплощено в системах, различных по своим физическим свойствам. Ментальные состояния являются не физическими, но функциональными отношениями. Следовательно, перенос интеллекта и сознания человека на кремниевую основу возможен. Эта идея стала водоразделом между философами, разделив их на противников и сторонников возможности отождествления сознания с деятельностью компьютера⁷⁸. По сути, функционализм Х. Патнэма явился логичным продолжением идей А. Тьюринга, интегрированным с идеями бихевиористского подхода к изучению человека, столь популярного в 60-е гг. прошлого века. Х. Патнэм считал, что в любой системе результат может определяться входными параметрами с определенной (допустимой) вероятностью, и потому называл любой живой организм «вероятностным автоматом»⁷⁹. Нынешние коннекционистские подходы к искусственному интеллекту, которые используют технологии нейронных сетей глубокого обучения — по сути используют подход Х. Патнэма в качестве парадигмы: «дайте нам более быстрые компьютеры и больше обучающих выборок, и мы повысим вероятность правильного предсказания!».

Однако машинный функционализм Х. Патнэма столкнулся со сложностью, которую философ так и не смог разрешить: объекты обладают качественными характеристиками, которые субъекты получают непосредственно во время наблюдения, то есть феноменология квалиа.

Ещё один мысленный эксперимент Х. Патнэма широко используется при обсуждении исследований по искусственному интеллекту и робототехнике: чтобы визуально представить проблему референции на объекты, он предложил картину «мозги в бочке», - наш мозг на самом деле

⁷⁸ Юлина Н. С. Очерки по современной философии сознания. М.: Канон+, 2015. 408 с.

⁷⁹ Putnam H. The nature of mental states // Putnam H. Philosophical papers. Vol. 2: Mind, language and reality. Cambridge: Cambridge University Press, 1975. P. 429–440. (Заметим, что позднее Патнэм отошел от этой позиции.)

находится в питательном растворе внутри специальной емкости, и в реальности не имеет контактов с окружающим миром, за исключением проводов, которые передают нам имитации различных сенсорных ощущений⁸⁰.

А. Тьюринг и Х. Патнэм проложили путь не только философскому функционализму, но и направлению исследований в области искусственного интеллекта, последователи которого утверждали, что в какой-то момент в будущем компьютерные программы могут превзойти интеллектуальные возможности человека — достаточно лишь написать программу, которая воспроизведёт входы и выходы в правильной последовательности.

Для конкретизации критики этого направления исследований Дж. Сёрл ввёл в 1980 г. само понятие «сильный искусственный интеллект», ведь чтобы опровергнуть существование какого-то объекта, его требуется определить хотя бы номинально.

Дж. Сёрл пошел по пути конструирования мысленных экспериментов, которые бы вскрыли значимость качеств, отсутствующих у тьюринговых машин, у современных ему роботов, — это смысл и интенциональность. Исследуя речевые акты при общении людей, Дж. Сёрл поставил в центр внимания психологию и философию сознания, нейробиологию и нейропсихологию.

В своей работе «Разумы, мозги, программы»⁸¹, Дж. Сёрл обрушился на физикализм и компьютерный функционализм А. Тьюринга и Х. Патнэма с позиций рассмотрения «смысла» и «намерения», развивая то направление в философии искусственного интеллекта, которое он назвал «биологическим натурализмом». Свою цель Дж. Сёрл в более поздней работе формулирует ещё более ясно и даже можно сказать дерзко: «я хочу забить последний гвоздь в

⁸⁰ *Putnam H. The Meaning of "Meaning" // Putnam H. Philosophical Papers. Vol. 1: Mind, Language and Reality. Cambridge: Cambridge University Press, 1975. P. 215–271.*

⁸¹ *Сёрл Дж. Разумы, мозги, программы // Тест Тьюринга. Роботы. Зомби / Пер. с англ. Д. Родионова; под ред. А. Ю. Алексеева. М.: МИЭМ, 2006. 120 с.*

гроб теории, согласно которой сознание является компьютерной программой»⁸².

Дж. Сёрл вполне разделяет оптимизм по поводу возможностей «слабого искусственного интеллекта»: на базе корректно описанных формальных правил (включая статистические методы, которые лежат в основе широко применяемых нейронных сетей) возможно моделирование среды или какой-либо ситуации.

Однако в ранее упомянутой статье «Разумы, мозги, программы» Дж. Сёрл формулирует аргумент, а, скорее, мысленный эксперимент, который получил название «китайской комнаты». Для сторонников функционализма и операционализма представляется очевидным, что компьютерные программы, получив достаточное количество информации, стандартных правил её использования и быстродействующую аппаратную базу, смогут моделировать «понимание» человека, имитируя осмысленность реплик в диалоге собеседников. Для опровержения концепции понимающих и сознающих компьютеров, которые оснащены только сенсорами, актуаторами и процессорами, Дж. Сёрл и конструирует следующий эксперимент⁸³: человек, не владеющий китайским языком, может имитировать его понимание, отвечая на вопросы с помощью списка правил по составлению иероглифов и набора китайских текстов с правильными ответами, ориентируясь на совпадение иероглифов, но не осознавая их значения.

Сегодня технически вполне реально обучить самые совершенные компьютеры всем сгенерированным текстам на всех языках мира (end-to-end deep learning), однако подобная весьма затратная система не будет проявлять понимания. Системы машинного перевода не могут работать с содержанием

⁸² Сёрл Дж. Открывая сознание заново / Пер. с англ. А. Ф. Грязнова. М.: Идея-Пресс, 2002. С. 5.

⁸³ Сёрл Дж. Разумы, мозги, программы // Тест Тьюринга. Роботы. Зомби / Пер. с англ. Д. Родионова; под ред. А. Ю. Алексеева. М.: МИЭМ, 2006. 120 с.

текстов, а потому ошибаются⁸⁴. Фактическая реализация мысленного эксперимент Дж. Сёрла показала ограниченность этого умпостроения.

Дж. Сёрл критиковал перенос сторонниками сильного искусственного интеллекта человеческой интенциональности на артефакт, которым, по его мнению, является компьютерная программа. Ведь программа даже в беспилотном автомобиле обладает понимание на уровне простейшего калькулятора: «понимание компьютера не является полным или частичным, оно отсутствует»⁸⁵. Чтобы понять китайский язык, недостаточно оперировать синтаксическими правилами (т. е. формальной логикой) или символами (семантикой). Необходима казуальность, причинная зависимость в связке мозг — интеллект — сознание, лишь тогда компьютеры в эксперименте с «китайской комнатой» поднимутся над уровнем формального моделирования последовательности возбуждения нейронов. Дело не в том, что когнитивные функции не могут быть представлены в виде вычислительных процессов и алгоритмов. Для Дж. Сёрла очевидно, что «вычислительные процессы и алгоритмы могут отлично существовать без каких-либо когнитивных функций»⁸⁶.

Интересно отметить, что когда речь идёт о понимании, то при этом герменевтическая традиция «используется разработчиками систем искусственного интеллекта двояким образом»⁸⁷, как для обоснования своего скепсиса в отношении создания Общего искусственного интеллекта, так и для углублённого понимания искусственного интеллекта естественных языков.

Однако искусственный интеллект, компьютер может быть не только объектом изучения сквозь оптику концепций человеческого сознания, но и

⁸⁴ Hofstadter D. The Shallowness of Google Translate // The Atlantic, 2018. URL: <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/> (дата обращения: 26.04.2020).

⁸⁵ Сёрл Дж. Разумы, мозги, программы // Тест Тьюринга. Роботы. Зомби / Пер. с англ. Д. Родионова; под ред. А. Ю. Алексеева. М.: МИЭМ, 2006. С. 11.

⁸⁶ Там же. С. 20.

⁸⁷ Шульга Е. Н. Когнитивная герменевтика. М.: ИФ РАН, 2002. С. 221.

наоборот, быть инструментом, полигоном для отработки новых подходов к изучению сознания человека, к проблеме «тела-сознания». Д. Деннет — один из наиболее ярких и последовательных представителей этого подхода.

Центральным пунктом рассуждений Д. Деннета становится утверждение о том, что самость-сознание — скорее иллюзия, создаваемая человеческой культурой. Оно не является каким-то спрятанным внутри каждого человека драгоценным феноменом. Скорее, это «виртуальная машина», специальный инструмент, который мы создаем на компьютерах, чтобы имитировать работу иного программного обеспечения. К примеру, несмотря на то что программа, написанная для компьютера с операционной системой Windows, не может быть запущена на компьютере, на котором работает MacOS, можно создать специальную виртуальную машину на MacOS, которая будет эмулировать Windows, делая возможным запуск программ этой операционной системы. Д. Деннет пытается обосновать утверждение, что наше сознание мало отличается от подобной «виртуальной машины», сходной с виртуальными машинами, которые создают пользователи компьютеров, желающие использовать различные операционные системы на одном и том же компьютере. С точки зрения Д. Деннета, «вполне правомерно утверждать, что соответствующим образом запрограммированный робот с компьютерным мозгом на кремниевой основе в принципе был бы сознательным и обладал бы самостью»⁸⁸.

Исходной базовой задачей науки об искусственном интеллекте, по мнению Д. Деннета, является «анализ когнитивных способностей личности путём представления её в качестве огромной сети субличностных специализированных элементов»⁸⁹.

⁸⁸ Юлина Н. С. Д. Деннет: самость как центр «нарративной гравитации» или почему возможны самостные компьютеры? // Вопросы философии. 2003. № 3. С. 104.

⁸⁹ Деннет Д. Насосы интуиции и другие инструменты мышления / Пер. с англ. З. Мамедьярова, Е. Фоменко. М.: АСТ, 2019. С. 113.

Сознание для Д. Деннета есть собрание различные самоорганизующихся роботов, каждый из которых приспособлен под свой круг задач (треков), в результате работы которых появляется множество набросков в разнообразных диспозициях — их определяют биологические факторы и социальные, среди которых можно выделить выученные паттерны поведения. При этом образ виртуальной машины, производящей феномен сознания, остаётся для Д. Деннета метафорой, причем гибкой, форма которой может меняться в разных книгах и статьях⁹⁰.

Но именно эта метафора — ключевая в описании Д. Деннетом феномена сознания: исследователи теоретической функциональной организации мультитрековой работы гомункулусов (или роботов) сознания должны сосредоточить внимание именно на уровне программного обеспечения работы мозга. Метафора позволяет соединить феноменологию сознания и физический мозг. Но как тогда выглядит архитектура сознания в представлении Д. Деннета? База — это физическая структура мозга и нейронная активность, они выступают аппаратным обеспечением сознания. Следующий уровень, как и с запуском программы на компьютере, это деятельность данной программы. Лишь информация, которую создает работающая программа, и есть сознание. То есть программный код, который можно зафиксировать на бумаге или на жестком диске, не является сознанием, но лишь виртуальным экземпляром программы, который ведёт свою деятельность, обрабатывая внешние сигналы и делая выводы.

Д. Деннет, развивая компьютерную метафору, заимствует из эволюционной теории Ч. Дарвина принцип «компетентности без понимания»: искусственный интеллект, компьютер может выполнять действия сколь угодно надёжно, но остается «слепым часовщиком»⁹¹, лишенным понимания.

⁹⁰ Юлина Н. С. Д. Деннет: самость как центр «нарративной гравитации» или почему возможны самостные компьютеры? // Вопросы философии. 2003. № 3. С. 104–120.

⁹¹ Докинз Р. Слепой часовщик / Пер. с англ. А. Гопко. М.: Corpus, 2015. 496 с.

При этом Д. Деннет рассматривает эволюцию искусственного интеллекта и компьютеров как чисто линейный, количественный рост: со времени изобретения универсальной машины Тьюринга, этого «универсального имитатора»⁹², компьютеры лишь наращивают вычислительную мощность.

Д. Деннет объединил концепцию эволюции по Ч. Дарвину и компьютерную метафору искусственного интеллекта в своей оригинальной теории функционализма, в «процессно-информационной модели искусственного интеллекта». То есть он попытался представить концепции сознания и искусственного интеллекта в рамках единой логичной структуры, проводя аналогию между «вычисляющими себя программами» и «мыслящими себя мыслями», причем компьютерный язык может быть предпочтительнее для описания мысленной деятельности, чем физические или ментальные коды.

Концептуальная история искусственного интеллекта в интерпретации представителей аналитической философии могла бы быть своего рода историей развития спектра физикалистско-функционалистских взглядов от радикальных физикалистов типа Р. Карнапа и Д. Армстронга до функционалистов типа Х. Патнэма и Д. Деннета, если бы не Т. Нагель, Дж. Сёрл, а за ними австралийский философ Д. Чалмерс, который специализируется на проблеме сознания, и в рамках современной философии искусственного интеллекта является наиболее ярким выразителем концепций дуализма⁹³.

Монография Д. Чалмерса «Сознающий ум. В поисках фундаментальной теории сознания»⁹⁴ — ключевая работа этого философа, в которой он предложил разделить общую проблему сознания на «трудную» и несколько

⁹² Деннет Д. Насосы интуиции и другие инструменты мышления / Пер. с англ. З. Мамедьярова, Е. Фоменко. М.: АСТ, 2019. С. 155.

⁹³ Юлина Н.С. Тайна сознания: альтернативные стратегии исследования // Юлина Н.С. Очерки по современной философии сознания. М.: Канон+, 2015. С. 179-221.

⁹⁴ Чалмерс Д. Сознающий ум. В поисках фундаментальной теории / Пер. с англ. В.В. Васильева. М.: УРСС: Книжный дом «Либроком», 2013. 512 с.

«легких». Суть «трудной проблемы сознания» в том, что текущая парадигма науки не в состоянии предложить путей её решения: отчего некоторые виды мозговой деятельности, связанные с переработкой информации, коррелируют с сознанием? В предельно острой формулировке этот вопрос звучит так: «В каком порядке должны быть расположены частицы, чтобы возникло сознание?»⁹⁵. Если же задача решается с помощью современных научных инструментов, например, путём нейробиологического поиска нейронных коррелятов сознания, то она «легкая».

Парадоксальность позиции Д. Чалмерса в том, что, оставаясь на позициях натуралистического дуализма, он отстаивает возможность создания сильного искусственного интеллекта и оппонирует как функционалистам (А. Тьюрингу, Х. Патнэму, Д. Деннету), так и биологическим натуралистам (Дж. Сёрлу, Т. Нагелю).

Д. Чалмерс разделяет свои возражения на две категории. Первую категорию он считает внешними «возражениями»⁹⁶. С помощью подобных контраргументов философы «...пытаются показать, что вычислительные системы не могут вести себя подобно когнитивным системам, так как люди обладают определёнными способностями, которые не могут быть реализованы технически»⁹⁷. Под описание таких возможностей подходят интуиция и творчество. Д. Чалмерс считает, что «внешние возражения наталкивались на трудности, связанные с успехами вычислительных симуляций физических процессов в целом»⁹⁸.

Критика Д. Чалмерсом «китайской комнаты» Дж. Сёрла сосредоточена на внутренних, когнитивных возражениях: имитирующие человека машины полностью лишены ментальной жизни, не имеют подлинного понимания,

⁹⁵ *Тегмарк М. Жизнь 3.0 / Пер. с англ. Д. Баюка. М.: Corpus, 2019. С. 430.*

⁹⁶ *Чалмерс Д. Сознательный ум. В поисках фундаментальной теории / Пер. с англ. В. В. Васильева. М.: УРСС: Книжный дом «Либроком», 2013. 512 с.*

⁹⁷ Там же. С. 358.

⁹⁸ Там же.

осознания объектов. Потому Д. Чалмерс скептически именует их «кремниевыми разновидностями зомби».

Д. Чалмерс утверждает, что с позиций дуализма основным аргументом в пользу сильного искусственного интеллекта является концепция имплементации. Имплементация — это связующее звено между физическими системами и абстрактными вычислительными объектами. Это перемычка, которая возникает в момент действия системы, включающей в себя обе составляющих. При создании математиками достаточно сложного, но правильного алгоритма, который сможет производить необходимые вычисления, появление такой перемычки, и, следовательно, «компьютерного сознания» неизбежно. Разумеется, в основе такого заявления о возможности сильного искусственного интеллекта лежит теория комбинаторных автоматов — отлично известная Д. Чалмерсу, как математику. Машину Тьюринга можно имплементировать, если имплементирован соответствующий комбинаторный автомат. Причем базой для комбинаторного автомата могут выступать и клеточные структуры (биологические автоматы), и язык программирования (условный C++) — все они позволяют обеспечить условия для вычислительной машины Тьюринга.

Д. Чалмерс посчитал, что Дж. Сёрл в вопросе о вычислениях поменял местами причину и следствие: вычисления будут имплементироваться практически любой предельно простой системой, даже состоящей из одного или двухэлементного комбинаторного автомата. Потому имплементация не лишает вычисления смысла, а напротив — нажатие кнопки RUN придаёт им смысл.

Надо отметить, что Д. Чалмерс в общефилософском плане занимает позицию панпсихизма, поэтому ему легко объяснить связь физического и психического, поскольку психическое для сторонников панпсихизма уже изначально содержится в физическом, является его неразвитой составляющей.

Несмотря на основательную критику физикалистских и радикально функционалистских интерпретаций искусственного интеллекта, инженеры и

исследователи искусственного интеллекта по-прежнему убеждены, что алгоритмы и программы сами по себе могут понять человека. И что нет необходимости специально привлекать для повышения результатов своей деятельности проблему сознания.

В 1960–70-е гг. в Советском Союзе философы широко обсуждали проблемы искусственного интеллекта. При Президиуме академии наук был создан и успешно функционировал Научный совет по кибернетике под председательством академика А. И. Берга. При совете работала философская секция (председатель Б. В. Бирюков), которая проводила научные конференции и симпозиумы по широкому спектру проблем кибернетики и искусственного интеллекта, осуществляла публикацию статей и монографий, привлекала к сотрудничеству ученых разных специальностей.

В отличие от концепций представителей аналитической философии, которые обсуждались выше, советские философы рассматривали проблему искусственного интеллекта в гораздо более широком плане, ставили в центр внимания актуальные вопросы кибернетики и самоорганизации, анализ природы информации, специфики и форм информационных процессов в их соотношении с физическими процессами и разработкой проблемы сознания, перспективы развития искусственного интеллекта (В. С. Тьютин⁹⁹, Б. В. Бирюков¹⁰⁰, И. Б. Новик¹⁰¹, Л. А. Петрушенко¹⁰², Д. И. Дубровский и др.¹⁰³), специально рассматривали вопросы робототехники (Е. И. Бойко¹⁰⁴). Эти исследования проводились в содружестве с выдающимися учёными, такими как А. И. Берг, А. Н. Колмогоров, Н. А. Бернштейн, П. К. Анохин, и

⁹⁹ Тьютин В. С. О природе образа: психическое отражение в свете идей кибернетики. М.: Высшая школа. 1963. 121 с.

¹⁰⁰ Бирюков Б. В. Кибернетика и методология науки. М.: Наука, 1964. 416 с.

¹⁰¹ Новик И. Б. Кибернетика. Философские и социологические проблемы. М., 1963. 208 с.; Новик И. Б. Философские вопросы моделирования психики. М.: Наука, 1969. 174 с.

¹⁰² Петрушенко Л. А. Самодвижение материи в свете кибернетики. Философский очерк взаимосвязи организации и дезорганизации в природе. М.: Наука, 1971. 292 с.

¹⁰³ Кибернетика. Мышление. Жизнь / Под ред. Б. В. Бирюкова. М., 1964. 512 с.

¹⁰⁴ Бойко Е. И. Сознание и роботы // Вопросы психологии. 1966. № 4. С. 169–177.

др., которые активно занимались ключевыми теоретическими и методологическими вопросами искусственного интеллекта и внесли в эту область большой вклад, полностью сохраняющий и ныне свое значение. Разработка советскими философами проблем искусственного интеллекта, проводимая в те годы, представлена в большом числе публикаций и требует специального историко-философского анализа.

В последние десятилетия западные философы, специализирующиеся на проблеме искусственного интеллекта, во многом устарились от осмысления конечной точки движения технологического развития искусственного интеллекта. Скорее ответ формулируется в рамках глобального эволюционизма — образ технологической сингулярности (В. Виндж¹⁰⁵). Последовательная рационализация, чисто количественное улучшение технологий, без опоры на ясную методологию или качественно новую философскую концепцию, не приблизит исследователей к решительным успехам в создании Общего искусственного интеллекта.

Последние 20 лет развития искусственного интеллекта и робототехники представляют собой поле битвы между двумя крупными парадигмальными подходами: классическим подходом, основанным на роли логико-символьных репрезентаций, оперирующих функциями сегментации и генерализации знаний, и коннекционистским подходом, основанным на симуляции нейронных связей человеческого мозга. Этот подход начался с работ А. Ньюелла и Г. Саймона в 60–70-х гг. прошлого века¹⁰⁶, которыми была сформулирована «Гипотеза физической символьной системы» (The Physical Symbol System Hypothesis): *«Физическая символьная система является необходимым и достаточным условием для произведения разумных действий».*

¹⁰⁵ Vinge V. The Technological Singularity. URL: <https://web.archive.org/web/20100621162528/http://www.kurzweilai.net/meme/frame.html?main=%2Farticles%2Fart0092.html> (дата обращения: 17.08.2020).

¹⁰⁶ Newell A., Simon H. A. Computer science as an empirical enquiry: Symbols and search // Communications of the ACM. 1976. Vol. 19. P. 113—126.

На основе этой гипотезы А. Ньювеллом и Г. Саймоном были разработаны первые когнитивные архитектуры в искусственном интеллекте, которые предполагали, что интеллектуальным роботам необходимо лишь проводить синтаксические действия над определённым набором символов (репрезентациями). Этот подход является одним из направлений машинного (компьютерного) функционализма. На основе данного подхода была разработана когнитивная архитектура SOAR. С эпистемологической точки зрения такие архитектуры отрицали роль телесности робота в познании мира и предлагали решать любые задачи за счёт увеличения количества репрезентаций и изучения связей между ними. Ярким проявлением такого подхода является база данных знаний об окружающем мире CYC.

Этот подход к реализации интеллектуальности роботов и машин получил серьёзную критику в работах Р. Брукса¹⁰⁷, А. Кларка¹⁰⁸, Х. Дрейфуса¹⁰⁹ и многих других авторов, которые отстаивали позицию связи познавательных функций интеллекта (как человеческого, так и машинного) и телесности, будучи убеждёнными, что классический взгляд машинного функционализма на роль репрезентаций в познании является «слишком церебральным» (Е. Спитцер)¹¹⁰. Кларк приводит пример исследования поведения самок сверчков, которые идентифицируют самцов сверчков с помощью уникальной системы локализации источника звука. Кларк утверждает, что этот процесс осуществляется полностью без внутренних репрезентаций окружающего мира, целиком полагаясь на механическое

¹⁰⁷ *Brooks R. A. Intelligence Without Representation // Artificial Intelligence. 1991. No. 47. P. 139—159.*

¹⁰⁸ *Clark A. Reasons, robots and the extended mind. // Mind & Language. 2001. Vol. 16(2). P. 121—145.*

¹⁰⁹ *Dreyfus H. L. What computers still can't do: A critique of artificial reason. Cambridge, MA: MIT Press, 1992. 408 p.*

¹¹⁰ *Spritzer E. Tacit Representations and Artificial Intelligence: Hidden Lessons from an Embodied Perspective on Cognition // Müller V.C. (ed.) Fundamental Issues of Artificial Intelligence. Switzerland: Springer International, 2016. P. 425–441.*

решение задачи самкой сверчка. Схожие механизмы для решения повседневных задач используются и людьми¹¹¹.

Аналогичный подход использовался Р. Бруксом в 90-х годах прошлого века в разработке бихевиористской архитектуры, которая привела к большому успеху в распространении роботов-пылесосов. Эта архитектура получила название «категоризированной архитектуры» («subsumption architecture») и широко используется для создания роботов, решающих однотипные задачи. Но у подобного подхода обнаружились сложности: робот, который хорошо решает одну задачу с помощью «мудрости своего тела», совершенно не способен к трансляции опыта решения задачи на схожую — робот, который научился ходить по камням с помощью такой архитектуры, не сможет эффективно огибать неожиданно возникающие препятствия.

С начала 90-х годов прошлого века начала свое развитие другая парадигмальная архитектура для интеллектуальной робототехники и искусственного интеллекта — мультиагентный подход, который «предоставил ещё один способ мышления об ИИ: создание агентов, которые могут эффективно действовать от имени субъекта»¹¹². Этот подход снова поднял вопрос об актуальности теста Тьюринга, изначальная цель которого была в том, чтобы подтвердить идею о создании машины, имеющей поведение, неотличимое от поведения людей. Но если мы хотим, чтобы агенты действовали от нашего имени и делали для нас всё возможное, то неважно, делают ли они тот же выбор, что и человек. Главной целью становится стратегия правильного выбора со стороны машины, а именно наилучшего из возможных. Таким образом, цель создания интеллектуальных роботов изменилась и перешла от создания агентов, которые делают выбор как человек, к агентам, которые делают оптимальный выбор для человека на

¹¹¹ Clark A. Reasons, robots and the extended mind. // *Mind & Language*. 2001. Vol. 16(2). P. 121—145.

¹¹² Wooldridge M. *The Road to Conscious Machines: The Story of AI*. London: Penguin, 2020. 388 p.

основе отрефлексированного и запрограммированного в робота множества ценностей.

Ценностно-ориентированный подход, основанный на максимизации полезности робота,— машины для человека, породил новую проблему, которая лежит на стыке робототехники и нейронаук: если роботы-агенты совершают оптимальный выбор и действуют в соответствии с ним, то человек может утратить способность к совершению такого выбора или ассоциированного с ним действия («используй способность, или потеряешь её»¹¹³).

Философско-методологические основания Общего искусственного интеллекта и робототехники ежегодно проходят серьёзное обсуждение на авторитетных мировых конференциях «Artificial General Intelligence» и «Robophilosophy»¹¹⁴.

В 2013 г. была выпущена фундаментальная работа «Alan Turing: His Work and Impact»¹¹⁵, посвящённая творческому наследию А. Тьюринга, в которую вошли как все его собственные работы, так и работы десятков исследователей в области искусственного интеллекта, формальной логики, аналитической философии и нейронаук.

В частности, на проведенной в 2015 г. конференции «Beyond Turing», организованной Г. Маркусом при участии таких исследователей искусственного интеллекта и робототехники как Б. Ленат, К. Форбус, С. Шейбер, Т. Подджио, Э. Мейрс, С. Адамс, Г. Банавар, М. Кэмпбелл, Ч. Орtiz, Л. Житник, А. Аграваль, С. Антоль, М. Митчелл, Х. Китано, В. Джарролд, Г. Маркус, О. Этциони и др. были рассмотрены новые типы

¹¹³ Use It or Lose It. URL: <https://neurosciencenews.com/brain-development-myelin-7224/> (дата обращения: 30.07.2020).

¹¹⁴ International Research Conference Robophilosophy 2020: Online conference, August 18-21, 2020. URL: <https://conferences.au.dk/robo-philosophy/> (дата обращения: 21.08.2020); The thirteenth annual conference on Artificial General Intelligence (AGI-20) take place Online, June 22-26. URL: <http://agi-conf.org/2020/call-for-papers/> (дата обращения: 29.07.2020).

¹¹⁵ Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. 944 p.

тестов для робототехники и искусственного интеллекта, сделаны обоснованные предложения по использованию воплощённости интеллекта для создания моделей Общего искусственного интеллекта. Кроме того, необходимо отметить отдельные работы и сборники при участии таких исследователей, как В. Нют, Дж. Копланд, Д. Бовен, М. Спревак, Р. Вилсон,¹¹⁶ Д. Дойч¹¹⁷, А. Кларк¹¹⁸, К. Варвик, Х. Ша, Х. Ишигуро, Б. Шнейдерман, С. Пенни и др.

Творческое наследие А. Тьюринга подвергается тщательному и критическому разбору в упомянутых выше работах. Проводятся специальные мероприятия, посвящённые исключительно прохождению теста Тьюринга различными способами, начиная от буквального воспроизведения игры в имитацию до робототехнических соревнований¹¹⁹.

Критическому анализу подвергаются не только работы А. Тьюринга, но и основы теорий всего исследовательского направления искусственного интеллекта. По выражению основателя направления структурного программирования Э. Дейкстра, «компьютерная наука (computer science) ровно столько изучает компьютеры, насколько астрономия изучает телескопы»¹²⁰.

Большие сомнения в возможности создания сильного искусственного интеллекта получили эмпирическую поддержку в 70-х годах прошлого века в связи с открытием так называемого эффекта «зловещей долины», согласно которому робот, имеющий значительное сходство с человеком, но далекий от

¹¹⁶ *Copeland J. Bowen J., Sprevak M., Wilson R.* The Turing Guide. Oxford: Oxford University Press, 2017. 544 p.

¹¹⁷ *Deutsch D.* Creative Blocks. URL: <https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence> (дата обращения: 18.07.2020).

¹¹⁸ *Clark A.* Can Philosophy contribute to an understanding of Artificial Intelligence? URL: <http://undercurrentphilosophy.com/medium/can-philosophy-contribute-to-an-understanding-of-artificial-intelligence/> (дата обращения: 02.08.2020).

¹¹⁹ *Brooks R.A.* The Case for Embodied Intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 500.

¹²⁰ *Gessler N.* The Computerman, The Cryptographer and The Physicist // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 526.

тождественности с ним, естественным образом отторгается нашим сознанием как субъект и не заслуживает такого же доверия, как при коммуникации с человеком (М. Мори, К. МакДорман, Н. Кадеки)¹²¹. Было достоверно установлено, что человек более позитивно воспринимает роботов, которые имеют гуманоидную форму, чем роботов, которые имеют облик манипулятора. Это различие между промышленными и гуманоидными роботами сохраняется за одним значимым исключением. Чем больше увеличивается антропоморфизм робота, тем большее отторжение вызывает робот у человека — робот входит в «зловещую долину». Это ставит очень значимый вопрос перед исследователями, которые пытаются применить методологию теста Тьюринга, используя робота вместо абстрактной вычислительной машины, которая имитирует разговор человека (тест кофемашины¹²², тест студента колледжа, тест устройства на работу¹²³ и др.), — является ли «зловещая долина» непреодолимым препятствием на пути увеличения похожести «машины» на человека в классическом понимании теста Тьюринга? Тест Тьюринга в классическом варианте становится бессмысленным, так как чем выше схожесть робота с человеком, тем большее отвращение может вызывать робот у человека в силу эффекта «зловещей долины».

Эпистемологические вопросы соотнесения искусственного интеллекта и робототехники подвергаются постоянному критическому разбору. Всё большую популярность приобретает концепция воплощённого интеллекта, которую обогащают парадигмы коннекционизма и логико-символьный подход. Однако в условиях быстроразвивающихся технологий робототехники

¹²¹ *Mori M.* The uncanny valley // *Energy*. 1970. Vol. 7(4). P. 33–35.

¹²² *Goertzel B.* Artificial general intelligence: Concept, state of the art, and future prospects // *Journal of Artificial General Intelligence*. 2014. Vol. 5. P. 1–46.

¹²³ *Goertzel B., Iklé M., Wigmore J.* The architecture of human-like general intelligence // *Wang P., Goertzel B.* Theoretical foundations of artificial general intelligence. Vol. 4. Amsterdam: Atlantis Press, 2012. P. 123–144.

и искусственного интеллекта невозможно определить лидирующую точку зрения, которая бы стала доминантой.

Применение теста Тьюринга для оценки уровня развития искусственного интеллекта предполагает качества, которые уподобляют его средству наблюдения, способу вопрошания в схеме научной рациональности, предложенной В. С. Степиным¹²⁴. Методологию Тьюринга можно соотнести с классической парадигмой: субъект-человек (судья) задаёт вопросы объекту-машине (в попытках отличить её от другого человека), следуя методу вопрошания «тест Тьюринга». А. Тьюринг полагал, что, максимально дистанцируясь от метода получения ответа, разделяя человека и машину стеной, преодолимой лишь для символической коммуникации, влияние средства вопрошания на результат теста можно сделать пренебрежимо малой величиной. В этом случае человек-исследователь может уловить наличие имитации чистого антропоморфного «мышления» в машине, какую бы физическую форму она не имела.

Объединение опыта последних семидесяти лет и новых технологических возможностей должно приводить к созданию новой методологии, которая позволит измерять уровень достижения искусственного интеллекта.

Во-первых, компьютер, обладающий Общим искусственным интеллектом, не должен быть строго антропоморфен. Созданная А. Тьюрингом стена, которая отделяет испытателя от испытуемого (человека или робота), фактически подталкивает людей-судей определять уровень адекватности искусственного интеллекта в сравнении с собой, со своей личностью. Это неизбежно порождает избыточный технологический антропоморфизм. Искусственный интеллект, общающийся с людьми и даже рассуждающий как человек, уже не является самым полным и достоверным критерием ответа на вопрос А. Тьюринга о мышлении машин.

¹²⁴ *Степин В. С.* Теоретическое знание. М.: Прогресс-Традиция, 2000. 744 с.

Во-вторых, необходимо рассуждать о доступном компьютерам разнообразии форм и методов познания. Искусственный интеллект должен использовать такие формы мыследеятельности как абстрагирование и конкретизация во всей их полноте. Тогда одним из критериев создания Общего ИИ ставится возможность самостоятельно формулировать понятия и моделировать картину мира. Сейчас процесс абстрагирования не доводится компьютером до требуемого уровня, и множество технологий используют только один из этапов перехода от феномена к понятию.

В-третьих, разнообразие форм общения, доступных человеку, должно реализовываться в компьютерных системах. При общении с людьми, осуществляемом посредством лишь символьных структур, компьютеры могут поддерживать определённый уровень адекватности, но моторика роботов всё ещё достаточно примитивна, как было отмечено ранее. Образ идеала для конструкторских разработок — это эмоционально окрашенное общение с помощью «пяти чувств». Однако пока он остается недостижимым.

В-четвёртых, искусственный интеллект, чтобы иметь возможность претендовать на статус наделённого субъектностью партнёра, пусть даже и младшего «друга» человека, неизбежно будет принимать участие в социальных практиках людей. Роль своеобразного помощника, который будет выполнять вторичные, периферийные функции в человеческой деятельности, учитывая потребности и эмоциональный фон людей, вполне доступен компьютерным системам с искусственным интеллектом.

Посттьюринговый подход в оценке искусственного интеллекта откроет возможности к созданию систем, обладающих качественно новым функционалом: они смогут вырабатывать разнообразные навыки в любых средах. Во-первых, осуществлять полный цикл формулировки понятий и свободно использовать новые понятия. Во-вторых, общаться с человеком, используя все пять чувств (учитывая как человеческие, так и пользуясь своими аналогами). В-третьих, менять структуру и облик системы, подстраиваясь под

очередные задачи, функции, трудовые операции. В-четвёртых, развивать социальную субъектность.

Глава 2. Роль тьюринговой методологии в исследованиях интеллектуальной робототехники

2.1. Определение значения тьюринговой методологии в контексте интеллектуальной робототехники

Ранее уже были рассмотрены вопросы, касающиеся роли идей функционализма в развитии искусственного интеллекта и робототехники, роль А. Тьюринга в их обосновании и использовании. Однако парадигма функционализма заслуживает более широкого взгляда с философских позиций, так как она ознаменовала фундаментальный поворот в формировании теоретических принципов научного познания. Дело в том, что на протяжении трёх столетий в науке господствовала парадигма физикализма, достигшая максимального влияния в начале XX века благодаря успехам физических наук. Она полагалась теоретической основой всякого подлинно научного объяснения, что получило широкое обоснование в рамках такого влиятельного философского направления в первой половине прошлого века, как логический позитивизм. Один из его ведущих представителей Р. Карнап формулировал это следующим образом: «все законы природы, включая те, которые управляют организмом, телом человека и человеческим обществом, вытекают из физических законов, необходимых для объяснения неорганических процессов»¹²⁵. Но это фактически означало, что всякое подлинно научное объяснение должно сводиться к физическому объяснению. Это в полной мере относится и к психологии. С позиции физикализма все виды и свойства ментальных феноменов, таких как чувства, мысли, идеи, наш многообразный субъективный опыт — всё это должно редуцироваться к физическим процессам, если исследователь хочет оставаться в рамках научного познания, а не «народной психологии». Другой ведущий представитель логического позитивизма К. Гемпель прямо заявляет:

¹²⁵ Carnap R. Discussion of Critics // The Philosophy of Rudolf Carnap / Ed. by P. A. Schillp. La Salle: OpenCourt, 1963. 883 p.

«Психология является составной частью физики»¹²⁶. Однако парадигма физикализма неизбежно заводит в тупик, когда предметом исследования становится сознание и социальные процессы.

В начале второй половины прошлого века в науке и философии сформировалась парадигма функционализма, в которой чётко обозначились альтернативы парадигме физикализма. Принцип изофункционализма систем стал их базовым отличием: *одна и та же функция может быть воспроизведена на разных по своим физическим свойствам носителях*. В эпистемологическом плане это выражается в логической независимости функциональных описаний от физических описаний, что позволяло рассматривать функциональное описание и объяснение в качестве адекватной теоретической основы для исследования биологических, психологических и социальных явлений, не отрицая ни в чём физические закономерности, но исключая радикальный физикалистский редукционизм. Вместе с этим парадигма функционализма, в формировании которой, как уже отмечалось, первостепенную роль сыграл именно А. Тьюринг, стала основой всего процесса развития искусственного интеллекта, создания компьютера и информационных технологий, ознаменовала новый, постиндустриальный этап нашей цивилизации.

Развитие искусственного интеллекта и робототехники внесло в общественное сознание новый взгляд на технику. Началось обсуждение вопроса: как нам сосуществовать с созданными нами техническими объектами, которые могут действовать на том же уровне интеллекта, что человек или животные. Обсуждение этого вопроса со временем расширяется, охватывая новые области научного знания. Две фундаментальные идеи, сформулированные с противоположных эпистемологических позиций, определяют весь современный дискурс, связанный с роботами или искусственным интеллектом.

¹²⁶ *Hempell K. G. The Logical Analysis of Psychology // Reading in Philosophical Analysis. N.Y.: Appleton-Century-Crofts, 1949. P. 378.*

Первая важнейшая концепция — «законы робототехники». Мысленно поместив себя на позицию робота, т.е. заняв эпистемологически позицию субъекта действия, А. Азимов сформулировал «три закона» робототехники, которые с точки зрения субъекта «я-робот» определяют законы, по которым живёт техника (на самом деле — любая техника) и определяют популярное мнение о том, как должны действовать роботы и искусственный интеллект по отношению к человеку. Сами законы робототехники А. Азимова многократно подвергались критическому анализу, и в диссертации нет смысла повторять сказанное. Однако важно отметить, что именно А. Азимов первым создал «я-методологию», т.е. метод поведения робота в мире, населённом создателями роботов, т.е. людьми.

Вторая концепция, получившая название «тест Тьюринга», направила всё развитие технологий искусственного интеллекта и робототехники по определённому руслу с помощью простого эпистемологического приёма: вопрос «может ли машина мыслить?» решался привнесением позиции наблюдателя во взаимодействие человека и машины и наделения человека статусом «судьи». Обе концепции, будучи «мысленными экспериментами» обрели не только научное значение для этики и эпистемологии искусственного интеллекта и робототехники, но и практически статус современной техномифологии, прочно прописавшись в научных статьях, научной фантастике и продуктах поп-культуры. В нашей работе мы исследуем вторую концепцию, которую определяем как тьюринговую методологию, и рассматриваем её в контексте развития интеллектуальной робототехники и искусственного интеллекта. Вопрос о том, можно ли предложенную А. Тьюрингом «игру в имитацию» назвать методологией обсуждался в философской литературе. В частности, С. Харнад¹²⁷ указал, что обе составляющих предложенного А. Тьюрингом названия «игра в имитацию»

¹²⁷ *Harnad S. Commentary on «Computing Machinery and Intelligence» // «Parsing The Turing Test» Philosophical and Methodological Issues in the Quest for the Thinking Computer / Ed. R. Epstein, G. Roberts, G. Beber. Amsterdam: Springer Netherlands, 2008. P. 27.*

являются ошибочными, направляющими исследователей по ложному пути. Слово «игра» создаёт ассоциацию с азартом или соревнованием, в котором обязательно есть выигравшая сторона, тогда как предложенный эксперимент является строго научным опытом «биологической обратной инженерной задачи». Слово «имитация» также содержит намёк на «фальсификацию», тогда как является строгой эмпирической методологией по проверке субъектности испытуемого.

А. Тьюринг сделал предположение о том, что мышление человека можно уподобить системе обработки абстрактных символов. Эта редукция стала базисом для тезиса А. Тьюринга об изоморфности мышления и вычисления: «Если считать результат труда вычислителей (т.е. людей, работающих над вычислениями) интеллектуальным, то почему нельзя сделать сходное предположение относительно машин, которые выполняют эти операции быстрее людей»¹²⁸.

Развивая предложенную аналогию, А. Тьюринг сформулировал образ «бумажной машины», которая в действительности просто человек, располагающий бумагой, карандашом и ластиком при выполнении стандартных операций. Подобная «машина», просто проводя последовательность операций, могла бы имитировать отдельные человеческие функции. В работе «Intelligent Machinery» 1948 г. «А. Тьюринг предложил сравнение человека и бумажной машины, способных осуществлять лишь символическую коммуникацию, поскольку находятся в разных комнатах и разделены стеной»¹²⁹. В ней А. Тьюринг изложил основы той концепции, которая два года спустя будет представлена как «игра в имитацию»¹³⁰ (работа «Computing machinery and intelligence»).

¹²⁸ Turing A. M. Intelligent Machinery // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 505.

¹²⁹ Там же. P. 516.

¹³⁰ Turing A. Computing machinery and intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 562–569.

Дополнительно в той же работе «Intelligent Machinery» А. Тьюрингом был проведён анализ возможностей и ограничений «воплощённого искусственного интеллекта» имитировать человека. Там А. Тьюринг сформулировал довод, что «отсутствие доступа этого создания к простым человеческим радостям, таким как еда, секс, спорт и многие другие, создаст препятствия для имитации интеллекта»¹³¹.

А. Тьюринг предполагал, что рациональнее сфокусировать усилия на имитации интеллектуальной деятельности человека в следующих областях:

1. Математика.
2. Криптография.
3. Лингвистические изыскания.
4. Переводы с одного языка на другой.
5. Разнообразные игры, такие как покер, бридж, шахматы, крестики-нолики.

А. Тьюринг выделял переводы с языка на язык, как наиболее полезные в утилитарном использовании искусственного интеллекта¹³². А выделение пяти указанных областей задавало вектор развития изучения искусственного интеллекта: постановка сравнительно однородных задач, частично решавшихся уже самыми первыми компьютерами 40-х гг., позволяла добиваться всё новых результатов простым наращиванием вычислительных мощностей. Возникла своеобразная инерция и даже замкнутый круг: исследователи стремились решать задачи в сравнительно узком секторе. Однако, как мышление людей, так и структуры социума сталкиваются с куда более широким набором «головоломок». Сектор применения компьютеров требуется расширять и, естественно, это расширение сопровождается завышенными ожиданиями. Но так же легко воспроизвести в других

¹³¹ *Turing A. M. Intelligent Machinery // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 508.*

¹³² Там же. P. 509.

прикладных областях те успехи, которые были достигнуты в криптографии, не удавалось. Это порождало, и закономерно продолжает порождать скептицизм.

Также А. Тьюринг обосновал важнейший тезис об эквивалентности мышления и вычисления и положил начало тому, что в дальнейшем стало машинным функционализмом, компьютеризмом и тестовым функционализмом¹³³. А. Тьюринг предложил инженерное, едва ли не праксеологическое решение задачи «Может ли машина мыслить?», — «игру в имитацию». Он не сосредотачивается на создании заведомо дискуссионных определений интеллекта машины и интеллекта человека. А. Тьюринг предложил просто сравнить их в «слепом» эксперименте, который был построен на возможности сопоставлять человеческое воображение с операциями вычислительных машин. На основе викторианской «игры в имитацию», которую А. Тьюринг преобразовал в сопоставление возможностей компьютера и человека вести диалог на естественном языке (то есть сформулировал «тест Тьюринга»), — и была построена методология создания искусственного интеллекта.

Важно подчеркнуть, что функциональные подходы, в разработку которых внёс большой вклад А. Тьюринг, открывали новые широкие перспективы в решении теоретико-методологических проблем не только искусственного интеллекта, но также нейронауки, психологических и социальных дисциплин, многих актуальных междисциплинарных проблем. Тем не менее, парадигма функционализма допускала различные интерпретации, часть из них носила бихевиористский или редукционистский характер, при которых исключалась роль проблемы сознания: феномены сознания полагались излишними для решения задач искусственного интеллекта или же отождествлялись с функциональными процессами, сводились к ним, как у некоторых представителей аналитической философии

¹³³ *Алексеев А. Ю.* Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга: диссертация на соискание ученой степени доктора философских наук: 09.00.08. М., 2015. С. 167.

(Д. Денет и др.). Такого рода интерпретация о «ненужности» сознания, близкая к бихевиористским установкам, была присуща также А. Тьюрингу. И заметим, что на первом этапе развития искусственного интеллекта она ему не препятствовала: задачи описываются в функциональных требованиях, на основе которых разрабатываются алгоритмы и программы, далее программа загружается в компьютер и задача решается, а результаты практически используются. Подобная методология сохраняется у большинства специалистов в области искусственного интеллекта до сих пор, поскольку для реализации практических, узкоспециальных приложений сознание не является необходимым элементом. Однако, создание Общего искусственного интеллекта и интеллектуальных роботов, основанных на его использовании, требует иных подходов. Сейчас мы стоим на пороге нового этапа развития искусственного интеллекта, и для осмысления его задач требуются новые методологические подходы и решения. Но для этого важно основательное философское осмысление предыдущего, тьюрингового этапа его развития, создавшего нашу цифровую, информационную цивилизацию.

После появления первых интеллектуальных роботов, имеющих кибернетическую обратную связь и строящих собственные динамически адаптируемые модели, операционалистский подход А. Тьюринга стал переноситься на робототехнику — «может ли робот выполнить любое действие не хуже человека?». Список типовых активностей «искусственного интеллекта», сформулированных Тьюрингом уже в начале развития искусственного интеллекта (игры, криптография, изучение языков, переводы, решение математических задач), существенно расширился. В результате применения этого подхода возникло большое семейство частных тестов Тьюринга, которые пытаются ответить на частные вопросы «может ли машина выполнить определённое действие?»¹³⁴. Такого рода частные тесты Тьюринга

¹³⁴ *Алексеев А. Ю.* Роль комплексного теста Тьюринга в развитии исследований искусственного интеллекта: доклад на заседании семинара НСММИ РАН «Философско-

слабо упорядочены, не имеют единой методологии и нуждаются в дополнительной классификации. Эта задача составляет важное условие разработки посттьюринговой методологии.

Исследователи искусственного интеллекта продолжили разрабатывать мысленные эксперименты, видя в них основу более точного, чем тест Тьюринга, критерия.

Такое развернутое описание дано, чтобы можно было раскрыть заданные современными условиями методологические сложности в оценивании «теста Тьюринга»:

1. Тест широко известен, причём в самой упрощённой формулировке: «Разговаривая по телефону с неизвестным абонентом, требуется разобраться за пять минут идёт общение с человеком или машиной».

2. Критерием научности теории выступает возможность её проверки. Инженерные исследования нуждаются в эмпирических, быстрых, простых и дешёвых тестах. Закономерно, что программное обеспечение, в образе которого чаще всего представляют ИИ, тестируется через общение с программой. Это утилитарное требование заложило инерцию восприятия «разумных машин».

Сформулируем три промежуточных вывода в анализе тьюринговой методологии:

Во-первых, пять областей исследования, изначально предложенных А. Тьюрингом, базируются исключительно на соотношениях символов и обменах символами. Шахматная партия, обмен репликами, взлом очередного шифра или доказательство теоремы — всё это существует в поле обработки символов.

Во-вторых, **стена**, разделяющая двух участников-людей, стала основой методологии А. Тьюринга (третий участник — компьютер). Стена позволяет

методологические и научно-теоретические проблемы искусственного интеллекта» (22 сентября 2015 г., Москва, Институт философии РАН). URL: https://iphras.ru/uplfile/ai/alekseev_presentatio.pdf (дата обращения: 19.06.2020).

редуцировать поток информации между компьютером и человеком до чисто символического уровня.

В-третьих, британский математик исходил из предположения, что проблема создания искусственного интеллекта состоит «главным образом в программировании». Бесконечное наращивание вычислительных мощностей, которое стало отличительной чертой компьютерной революции, не рассматривалось А. Тьюрингом как необходимое условие для убедительной победы при «игре в имитацию». Объем компьютерной памяти, который А. Тьюринг считал граничным, достаточным для этой игры — 10^9 бит — был достигнут уже в конце 80-х гг. прошлого века. Но в те годы значимых прорывов в создании Общего искусственного интеллекта не случилось. Более того, именно этот период совпал с периодом разочарования и скепсиса по отношению к технологиям искусственного интеллекта, так называемой «второй зимой» (1987–1993 гг.).

Это показывает, что чисто количественное наращивание вычислительных мощностей не позволяет машинам учитывать подтексты человеческого общения. Первоначально на необходимость использовать многоуровневость сообщений указал А. Тьюринг, отмечая проблемность задачи перевода: «...она кажется наиболее зависимой от осуществимости работы органов чувств и движения»¹³⁵. Однако данное замечание британского математика практически игнорировалось несколькими поколениями исследователей, для которых показателем интеллекта было адекватное лингвистическое поведение и способность соревноваться с игроками-людьми. Однако, великолепная в частностях, эта методология оказалась недостаточной для выполнения всех задач естественного интеллекта в рамках одной системы.

Большое семейство разнообразных тестов на искусственный интеллект было создано на основе методологии А. Тьюринга. Практически все они направлены на решение узких, частных задач. Ниже эта совокупность тестов

¹³⁵ *Turing A. Digital Computers Applied to Games // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 626–644.*

будет рассмотрена, и в ней будут вскрыты предпосылки для лишения теста Тьюринга статуса критерия в вопросе создания полноценного искусственного интеллекта.

2.2. Методологическая роль классификации частных тестов Тьюринга в развитии интеллектуальной робототехники

Несовершенство предложенного А. Тьюрингом подхода сразу же дало обильную почву для модернизации первоначальной идеи — сопоставления машины и человека через различные смысловые или деятельностные практики — начиная от решения творческих задач, до выявления психологических отклонений у тестируемых. Модернизацию можно проводить за счёт двух способов. Во-первых, можно изменять методику тестирования, сохраняя область тестирования. Во-вторых, можно применить методику тестирования к другой области. По сути, комбинацией различных параметров и областей тестирования можно создать практически неограниченный список различных «тестов». Например, может ли робот собирать грибы лучше, чем человек?

А. Ю. Алексеев ввел понятие частного теста Тьюринга: на основе тьюринговой методологии устанавливается степень «когнитивной компетентности систем различной природы — человеческих, биологических, технических, социальных»¹³⁶. Частный тест Тьюринга в определенном смысле является противоположностью комплексному тесту Тьюринга, понятие которого также введено А. Ю. Алексеевым. Различие между ними кардинальное — частный тест Тьюринга проверяет какую-либо одну когнитивную способность машины, а комплексный тест Тьюринга призван ответить на вопрос «может ли машина всё?». Одновременно, А. Ю. Алексеев признаёт за частным тестом Тьюринга функцию определения конечной точки создания программно-аппаратного комплекса, так как если тестирование

¹³⁶ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛЛ, 2013. С. 95.

показывает, что машина выполняет тест успешно, то она далее может успешно выполнять реализованную функцию.

Несмотря на большое число попыток усовершенствовать или модернизировать тест Тьюринга за последние семьдесят лет, нами не было обнаружено серьёзных попыток провести систематизацию всех известных тестов Тьюринга. Единственное предложение по категоризации всевозможных тестов на когнитивные способности машины было сделано А. Ю. Алексеевым, предложившим разделить «частные тесты» на совершенные, т.е. полностью удовлетворяющие условиям А. Тьюринга, «воспроизводящие все ключевые идеи оригинального теста Тьюринга и подстраивая их к собственным целям и задачам»¹³⁷ и «несовершенные», удовлетворяющие условиям А. Тьюринга лишь частично, «акцентирующие внимание как минимум на одной компоненте» без «целостного охвата»¹³⁸. Согласно А. Ю. Алексееву, сам оригинальный тест Тьюринга относится к совершенному виду тестов. Несмотря на то, что предложенная классификация является вполне приемлемой, она не даёт возможностей для создания новых путей развития искусственного интеллекта и робототехники, по сути дела ограничиваясь лишь оценкой того, насколько предложенная методика тестирования соответствует тому, что предложил А. Тьюринг семьдесят лет назад. Во многом это совпадает с подходом, обозначенным Р. Курцвейлом¹³⁹, согласно которому прогресс технологий по отношению к достижению цели создания Общего искусственного интеллекта можно измерить по количеству выполненных частных тестов Тьюринга.

Для дальнейшего рассуждения и попытки классификации необходимо рассмотреть несколько видов частных тестов Тьюринга (некоторые были рассмотрены А. Ю. Алексеевым, а некоторые анализируются впервые). При

¹³⁷ Там же. С. 96.

¹³⁸ Там же.

¹³⁹ Kurzweil R. The age of spiritual machines: When computers exceed human intelligence. New York: Penguin, 2000. P. 270–280.

их рассмотрении описываются условия применения теста, определяется критерий успешного прохождения и возможные пути усовершенствования.

Рассмотрим сложившийся исторически набор частных тестов Тьюринга:

Исходный, канонический тест Тьюринга. Его подробный анализ содержится в ряде отечественных и зарубежных работ (к примеру, А. Ю. Алексеев¹⁴⁰). Тест имеет трёх участников: человек (Ч), компьютер (К) и судья (С). С, К и Ч разделены стенами, которые прозрачны только для символической коммуникации. Задача Ч и К убедить С в том, что каждый из них является человеком. Если С будет ошибаться одинаково часто в определении того, где находится К (в сравнении с игрой, где участвует только Ч), то это означает, что интеллектуальные способности К и Ч одинаковы.

Параноидальный тест Колби. Сформулирован в статье «Искусственная паранойя»¹⁴¹. Предложена схема несовершенного теста Тьюринга: судья — это врач-психиатр, в роли человека параноик, а компьютер имитирует параноика. Всё взаимодействие Ч, К и С ведётся через терминал в виртуальной среде с помощью языкового интерфейса. Опыт тестирования¹⁴² К. Колби показал: психиатры не в состоянии различить слова компьютера и человека практически в половине случаев. По мнению А. Ю. Алексеева основанием для прохождения компьютером теста Колби стали редуцированность, лингвистическая недостаточность и просто хаотичность ответов психически больных лиц. Распространённые сейчас программы генерации текстов, особенно имитирующие научные работы, в состоянии проходить тест Колби.

Тест леди Лавлейс. Имя Ады Лавлейс, которая жила в XIX веке, была помощницей Ч. Бэббиджа и считается первой программисткой, и которую А. Тьюринг упомянул в работе 1950 г., рассуждая о «возражении Лавлейс», —

¹⁴⁰ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИнтелЛЛ, 2013. С. 55.

¹⁴¹ Colby K. M., Hilf F. D., Weber S. Artificial Paranoia // Artificial Intelligence. 1971. Vol. 2. P. 1–25.

¹⁴² Там же.

стало основой исторической аллюзии, давшей имя тесту. П. Белло, С. Брингсйорд, Д. Феруччи в работе «Творчество, тест Тьюринга и (улучшенный) тест Лавлейс» сформулировали следующую концепцию: «искусственный агент К, созданный человеком-разработчиком (Р), проходит тест Лавлейс тогда и только тогда, когда: 1) К создаёт на выходе некий артефакт (А) (с учётом того, что А. Лавлейс была дочерью Лорда Байрона, предполагалось, что А — поэма); 2) А — это не результат 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.»¹⁴³. Три исследователя не ограничились перечисленными условиями, потребовались дополнительные: 1) было сделано допущение, что подтасовки и жульничество, якобы, не помогут помочь пройти тест Лавлейс; 2) уже созданные на момент формулировки теста программы не могут заниматься творчеством и не смогут пройти тест, потому что специфика их умозаключений известна разработчикам теста, что исключает возможность стилизации, рерайтинга и т. п.; 3) компьютер должен иметь возможность автономно модифицировать, дополнять свою программу, то есть компьютеру должно быть доступно самообучение.

Схема Винограда в оценке здравого смысла. Данный тест не является напрямую диалоговым тестом и его можно отнести к несовершенным частным тестам Тьюринга. Однако он является очень важным для проверки насколько системы искусственного интеллекта могут опираться на здравый смысл, очевидный людям. Схема Винограда основана на предложении Г. Левеска в 2011 г. в качестве альтернативы тесту Тьюринга¹⁴⁴. Главной его особенностью является простой формат вопросов, который может охватывать многие

¹⁴³ *Алексеев А. Ю.* Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИнтеЛЛ, 2013. С. 76.

¹⁴⁴ *Levesque H., Davi E. and Morgenstern L.* The Winograd Schema Challenge. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference (KR2012)*. Palo Alto, CA: AAAI Press, 2012. P. 552–561.

области здравого смысла. Вопросы выбираются таким образом, чтобы они не требовали специальных знаний или подготовки и на них было легко ответить людям. Например: «Приз не влезает в коричневый чемодан, потому что он слишком большой. Что имеется в виду под «он»? или «Джим успокаивал Кевина, потому что тот был расстроен. Кто был расстроен?»¹⁴⁵.

Инвертированный тест Ватта. С. Ватт высказал предположение¹⁴⁶, что из-за явления психики, известного как наивно-психологический антропоморфизм, тест Тьюринга могут пройти программы, не обладающие свойствами интеллекта, — люди приписывают свои психические свойства любым системам — органическим и неорганическим. Машины, равно как здания и статуи, не стали каким-то исключительным случаем. Чтобы преодолеть этот стереотип мышления, необходимо поменять местами игроков. Оценивать надо не ответы, как таковые, а способность выносить правильные суждения о сути таких ответов. Судья должен быть судим. Внешне частный тест Тьюринга проводится в неизменном виде, но судья-человек должен сопоставлять ответы «компьютера» и «человека» с решениями компьютеризированного арбитра. Если судья-компьютер начинает приписывать интеллект тестируемой системе, которая им не обладает, то интуитивно понятен вывод: судья-компьютер есть машина, а не человек. Очевидно, что данный тест предполагает манипуляцию символьной информацией в виртуальном пространстве компьютерных терминалов (в силу сходства с оригинальным тестом Тьюринга). Но этот же тест поднимает проблему «наблюдателя сложности» — при дальнейшем развитии техники человек-судья помещается в специфические условия: «мир сложности становится миром сложных событий, развертывающихся в мире «черных

¹⁴⁵ *Ackerman E.* Can Winograd Schemas Replace Turing Test for Defining Human-Level AI? // IEEE Spectrum. URL: <https://spectrum.ieee.org/automaton/artificial-intelligence/machine-learning/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence> (дата обращения: 19.07.2020).

¹⁴⁶ *Watt S.* Naive Psychology and the Inverted Turing Test // Psychology. 1996. No. 7 (14). URL: <http://psycprints.ecs.soton.ac.uk/archive/00000506> (дата обращения: 03.08.2020).

лебедей» (Н. Талеб), так называемых «редких» событий, которые становятся, однако, всё более вероятными»¹⁴⁷. Разница в скорости обработки информации человеком и машиной уже составляет три порядка и может возрасть: внутреннее время судьи-человека принципиально отличается от внутреннего времени компьютера, и за счёт этой разницы возможен анализ поведения человека со стороны техники, организация псевдослучайных совпадений и т. п.

Китайская комната Сёрла. Описание мысленного эксперимента Дж. Сёрла было сделано ранее в Главе 1. Однако здесь отметим, что А. Ю. Алексеев считает¹⁴⁸ этот мысленный эксперимент частным тестом Тьюринга, который не может быть пройден в принципе. По мнению А. Ю. Алексеева, роли человека и компьютера выполняет сам автор мысленного эксперимента (Сёрл-в-комнате/Searle-in-the-room). По сути, ситуация мысленного эксперимента мыслится от первого лица, причём реальных судей — два человека: англичанин и китаец.

Психофункциональный тест Блока. Н. Блок скептически высказывается о классическом тесте Тьюринга, «показывая, что «игра в имитацию» является разновидностью оценки поведения в контексте лингвистического бихевиоризма»¹⁴⁹ (работа «Психологизм и бихевиоризм»). Он предлагает проверять «способность продуцирования осмысленной последовательности вербальных реакций на некоторую последовательность вербальных стимулов, но не акт продуцирования»¹⁵⁰. Если создать машину, которая сможет давать осмысленную последовательность вербальных реакций на вербальные стимулы, то она сможет пройти тест. Опосредованно ответы

¹⁴⁷ Аршинов В. И. Конвергентные технологии в контексте постнеклассической парадигмы сложности // Сложность, разум, постнеклассика. 2015. № 3. С. 52.

¹⁴⁸ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛ, 2013. С. 80.

¹⁴⁹ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛ, 2013. С. 78.

¹⁵⁰ Там же. С. 78.

этой машины заданы инженерами и программистами. Именно люди комплектуют наборы правильных ответов для суммы вопросов-стимулов. При внешней оценке поведения такой машины, она будет выглядеть интеллектуальной. Но при вскрытии коробки системного блока становится понятно, что её правильные ответы — это лишь копирование заданных ответов программистов, своим интеллектом она не обладает, а как попугай воспроизводит какие-то данные. Если библиотека правильных ответов будет достаточно велика, то машина сможет ответить на любой вопрос судьи. Единственный способ различить компьютер и человека можно назвать «анатомическим»: когда вскрыются внутренности устройства реализации, станет ясно, что это устройство не адекватно сумме вопросов, ответы — в памяти.

Тест Френча на когнитивные способности. Сформулирован сравнительно поздно, уже в 1990 г. Отправной точкой рассуждений Д. Френча стало утверждение, что тест Тьюринга: «умело обошёл бескрайнее философское болото проблемы дух/тело, однако он не способен исследовать глубинные, подсознательные области человеческого интеллекта. Поэтому в тест Тьюринга следует включить так называемые субкогнитивные вопросы, т.е. любые вопросы, позволяющие идентифицировать низкоуровневую когнитивную структуру, например, подсознательную сеть ассоциаций»¹⁵¹. Каждый человек за свою жизнь увидел огромное, практически бесконечное количество образов. Все они могут быть актуализированы с помощью сети ассоциаций и с помощью аналогий. Общество, культурный багаж и вся биография индивида-игрока определяют его ассоциативные ансамбли. То есть, для прохождения теста Френча — он описан в работе «Субкогнитивные способности и границы теста Тьюринга» — надо воспринимать окружающий мир, как человек, и жить, как человек. Но тогда получается, что тест Френча

¹⁵¹ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛЛ, 2013. С. 77.

выявляет скорее социокультурную обусловленность человеческого интеллекта, чем интеллект «вообще».

Тест Харнада. «Разумы, машины и Тьюринг: неразличимость неразличимости» (S. Harnad, 2001). Суть теста: использовать многоуровневую классификацию тестов Тьюринга для решения различных задач и установления разного уровня интеллекта. Выделяется пять уровней тестирования в зависимости от сложности тестируемой системы:

- T1 — неразличимость Ч (человек) и К (компьютер) в локальной задаче, например, настольной игре. Пример: AlphaGo;
- T2 — неразличимость Ч и К в вербальном диалоге (переписка по почте), классический тест Тьюринга. Пример: чат-боты;
- T3 — неразличимость на уровне диалогов и сенсорно-моторных действий робота, манипулирующего предметами окружающего мира. Согласно С. Харнаду, именно этот уровень является подлинным тестом Тьюринга. Пример: роботы из вселенной фильмов о терминаторах (роботы моделей T-800/T-1000);
- T4 — неразличимость на уровне внешних действий и внутренних структур, функций робота, включая нервную систему (искусственный сапиенс). Пример: репликаны из романа Ф. Дика «Снятся ли андроидам электрические овцы»;
- T5 — искусственный человек, неотличимый от естественного ни в чём, включая атомы и молекулы. Пример: «подкидыши» из романа А. и Б. Стругацких «Жук в муравейнике»¹⁵².

¹⁵² Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛЛ, 2013. С. 82–83.

Комплексный тест Тьюринга-Алексеева. А. Ю. Алексеевым в 2006 г. выдвинуто предложение объединить все частные тесты Тьюринга в один, комплексный тест, который тестирует весь спектр когнитивных возможностей человека за счёт совмещения коннекционистской и символической парадигм искусственного интеллекта на базе машины С. Н. Корсакова¹⁵³. В этом тесте проверяется возможность имитации машиной когнитивного поведения, схожего с человеком, в искусственных мирах. По сути, комплексный тест Тьюринга-Алексеева — это симуляция виртуального мира, в котором виртуальный сапиенс проходит тестирование во всех аспектах когнитивного взаимодействия с симуляцией. По своей природе данный тест является виртуальным, хотя и может проходить как в символической форме, так и в форме проявления познавательных способностей искусственным интеллектом, погруженным в симуляцию.

Тесты Брукса. Р. Брукс в 2018 г. предложил¹⁵⁴ ряд новых тестов, имеющих более комплексную направленность, чем классический тест Тьюринга и ряд других тестов, рассмотренных выше. Р. Брукс предложил рассматривать способности ребёнка как индикатор технологического достижения в области искусственного интеллекта и робототехники, уходя от тьюринговой парадигмы «разговорного» искусственного интеллекта и человека, общающихся сквозь стену. Можно использовать компетентностный подход, суть которого излагается в тезисах: 1) Робот должен обладать возможностью распознавать окружающие его объекты. При этом, как минимум, он должен достигать уровня адекватности двухлетнего ребенка. 2) Уровень понимания естественной речи, характерный хотя бы для четырехлетнего ребенка, — второй необходимый навык. 3) Шестилетний ребенок выступает эталоном для уровня ловкости рук и тонкой моторики. 4)

¹⁵³ Речь идет о «машине для сравнения идей», концепцию которой выдвинул еще в XIX веке С. Н. Корсаков. (Алексеев А. Ю. Тест Тьюринга: 66 лет спустя // Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (17–18 марта 2016 г., г. Москва, философский факультет МГУ им. М. В. Ломоносова). М.: ИИнтелл, 2017. С. 154–189.)

¹⁵⁴ Brooks R. A. Steps Toward Super Intelligence IV, Things to Work on Now // RODNEY BROOKS Robots, AI, and other stuff. Rodney Brooks., 15 July 2018 г. URL: <https://rodneybrooks.com/forai-steps-toward-super-intelligence-iv-things-to-work-on-now/>. (дата обращения: 08.08.2020).

Наконец, навыки социальной коммуникации не ниже уровня восьмилетнего ребенка совершенно необходимы роботам. Данный комплексный тест в силу своей направленности проходит через физическое или символическое взаимодействие и может быть реализован как виртуально, так и физически.

Тест Э.ЛЕНА. В Лаборатории робототехники Сбербанка в 2019 г. была разработана и запатентована специализированная платформа, позволяющая преобразовать текст в видеоизображение теледиктора¹⁵⁵. Платформа названа «Э.ЛЕНА — Электронная Лена». Утверждение, что визуализация искусственного интеллекта имеет большое значение, раньше служило аргументом скорее в научно-популярной или просто в популярной литературе. Исследователи не рассматривали визуализацию искусственного интеллекта в качестве объекта исследования в силу того, что не существовало самой технологии. Однако сейчас такая технология появилась и в скором времени станет широко доступна. Предлагается проводить тест на восприятие цифрового теледиктора путем сравнения его с диктором-человеком¹⁵⁶. Совершенно логично, если такой тест располагается в области невербального взаимодействия в виртуальном мире.

Тест ImageNet. Ф. Ли предложила использовать масштабную базу данных изображений ImageNet для улучшения качества распознавания образов в 2006 г. На текущий момент в базе содержится более десяти миллионов изображений, которые прошли ручную аннотацию, включая точные координаты предметов, которые есть на этих изображениях. С 2010 г. на этой базе данных проводится соревнование по компьютерному зрению ILSVRC¹⁵⁷ (англ. ImageNet Large Scale Visual Recognition Challenge — кампания по широкомасштабному распознаванию образов в ImageNet). Создатели новых программ сравнивают их возможности по распознаванию объектов и сцен, взятых из ImageNet. В 2012 г. произошел прорыв благодаря

¹⁵⁵ *Ефимов А. Р., Гонноченко А. С., Владимиров М. А.* Способ и система для создания мимики на основе текста // Патент на изобретение RU 2 723 454. 2020. С. 1.

¹⁵⁶ *Ефимов А. Р.* Технологические предпосылки неразличимости человека и его компьютерной имитации // Искусственные общества. 2019. Т. 14. С. 74–80. DOI: 10.18254/S207751800007645-8. URL: <https://artsoc.jes.su/s207751800007645-8-1/> (дата обращения: 16.06.2020).

¹⁵⁷ Large Scale Visual Recognition Challenge (ILSVRC). URL: <http://www.image-net.org/challenges/LSVRC/> (дата обращения: 20.07.2020).

использованию нейронных сетей глубокого обучения. В настоящий момент распознавание образов на этой базе превзошло человеческие возможности. Это подобно тому, как если бы диалоги в классическом тесте Тьюринга строились только вокруг одной темы (например, погоды).

Игровые тесты. Возможности игры для тестирования уровня искусственного интеллекта были замечены и описаны еще А. Тьюрингом в начале 50-х годов прошлого века. Потребовалось более сорока лет, чтобы компьютеры смогли научиться играть в шахматы лучше, чем люди, и почти семьдесят лет, чтобы они достигли уровня, превосходящего сильнейших игроков в го, которая считается наиболее сложной из игр. Однако, потенциал игр не ограничивается лишь настольными играми. Компьютерные игры в виртуальной симуляции также являются важнейшим аспектом «тестового функционализма» по А. Ю. Алексееву. В частности, компании Google (DeepMind) и OpenAI активно используют различные типы компьютерных игр для испытаний различных нейросетевых моделей, включая обучение с подкреплением и обучение без учителя¹⁵⁸.

Тесты NIST. Голос и звук в целом являются важной частью когнитивного инструментария человека. Поэтому Национальный институт стандартов и технологий США проводит постоянный конкурс лучших систем распознавания и синтеза голоса¹⁵⁹. Данный тест можно отнести к частичным, несовершенным тьюринговым тестам, так как в нём тестируется возможность компьютера по распознаванию голоса в сравнении с человеком.

Тест Ишигуро. Известный японский исследователь-робототехник Х. Ишигуро предложил «тотальный» тест Тьюринга¹⁶⁰. В данном тесте С должен выбирать не на основе символьного диалога с машиной, но на основе личного, физического опыта аудиовизуального взаимодействия с роботом-андридом. В случае, если С не может отличить с кем происходит взаимодействие, то тест Ишигуро считается пройденным. Данный тест по

¹⁵⁸ Парфёнов В. ИИ победил всех в Dota2. Сразиться с ним теперь может каждый // Популярная механика. URL: <https://www.popmech.ru/technologies/news-475872-ii-pobedil-vseh-v-dota2-srazitsya-s-nim-teper-mozhet-kazhdy/> (дата обращения: 05.08.2020).

¹⁵⁹ NIST 2019 Speaker Recognition Evaluation. URL: <https://www.nist.gov/itl/iad/mig/nist-2019-speaker-recognition-evaluation> (дата обращения: 20.07.2020).

¹⁶⁰ Ishiguro H. Android Science // Kasaki M, Ishiguro H, Asada M. et al. Cognitive Neuroscience Robotics. Springer, 2016. P. 193–234.

определению происходит в физическом пространстве, однако сам робот может находиться под управлением оператора. Следовательно, когнитивные возможности робота не исследуются в данном тесте, но исследуется восприятие робота человеком с помощью невербальных сигналов.

Тест Зимнего города. Одно из наиболее важных и приоритетных направлений современной робототехники — создание беспилотных транспортных средств. Робот-такси станет возможным тогда, когда будет доказано (с помощью тестов), что автомобиль под управлением искусственного интеллекта является более безопасным транспортным средством, чем под управлением человека. Этому аспекту развития технологий искусственного интеллекта и робототехники был посвящён конкурс «Зимний город», который прошёл в России в 2019 г. Суть конкурса в том, что автомобиль под управлением робота должен проехать по маршруту, не совершив ни одного ДТП и выдержав среднюю скорость движения в зимних условиях (25 км/ч). Данный тест был разработан коллективом авторов под руководством диссертанта¹⁶¹.

Тесты роботов DARPA. Агентство Передовых Оборонных Исследований США проводит различные конкурсы, направленные на создание новых технологий робототехники и искусственного интеллекта. Один из таких конкурсов — DARPA Robotics Challenge прошёл в США в 2015 г. и был посвящён имитации действий человека роботами в условиях ликвидации стихийного бедствия¹⁶². Конкурс включал в себя сложные манипуляционные и когнитивные задачи, которые по определению происходили в реальном физическом мире, требовали высокой точности распознавания объектов мира, однако не нуждались в символической, вербальной или невербальной коммуникации с человеком.

Выполненная выше аналитическая работа по рассмотрению далеко не полного перечня частных тестов Тьюринга помогает выяснить некоторые общие черты, которые позволяют предложить методологическую

¹⁶¹ Фонд «Сколково», АСИ и РВК объявили о запуске конкурсов НТИ. URL: <https://upgreat.one/media-center/smi/fond-skolkovo-asi-i-rvk-obyavili-o-zapuske-konkursov-nti/> (дата обращения: 07.08.2020).

¹⁶² DARPA Robotics Challenge. URL: https://en.wikipedia.org/wiki/DARPA_Robotics_Challenge (дата обращения: 23.07.2020).

классификацию. Прежде всего, можно заметить, что значительное число частных тестов Тьюринга, особенно предложенных до 2008 года, опираются на символьное взаимодействие в виртуальном мире. Таковы тесты Колби, Сёрла, схема Винограда, тест Френча, тест Ватта, тест Лавлейс и классический тест Тьюринга. В этих тестах весь информационный обмен между судьями, человеком и компьютером (роботом) выполняется с помощью символов, которые тестируемые и тестирующие должны печатать на компьютере (или ином устройстве). Об этом говорили многие исследователи наследия А. Тьюринга, включая А. Ю. Алексеева, К. Варвик, Х. Ша и др. Однако типология «виртуальных» тестов и сравнительный анализ с другими типами тестов, основанных на невербальных коммуникациях в физическом мире, ускользала ранее от внимания исследователей Общего искусственного интеллекта. Диссертантом была предложена методологическая классификация тестов Тьюринга в соответствии с двумя измерениями¹⁶³ (осями): 1) ось от символического (вербального) к невербальному взаимодействию человека как субъекта с роботом (или искусственным интеллектом) как объектом, вторая — проходит от виртуального к физическому взаимодействию человека с искусственным интеллектом, или вообще с любым интеллектуальным роботом.

На рисунке 1 частные тесты расположены по их отношению к предложенным выше осям *вербальное-невербальное* и *виртуальное-физическое*. Эти оси образуют двумерную систему координат. Предлагается назвать это общее двумерное пространство, в котором располагаются все мыслимые частные тесты Тьюринга — «пространством тьюрингоподобных тестов». В нем можно выделить четыре области. Рассмотрим их подробнее.

¹⁶³ Ефимов А. Р. Посттьюринговая методология: разрушение стены на пути к общему искусственному интеллекту // Интеллект. Инновации. Инвестиции. 2020. № 2. С. 74–80. DOI: 10.25198/2077-7175-2020-2-74. URL: http://intellekt-izdanie.osu.ru/arch/2020_2_74.pdf (дата обращения: 28.06.2020).

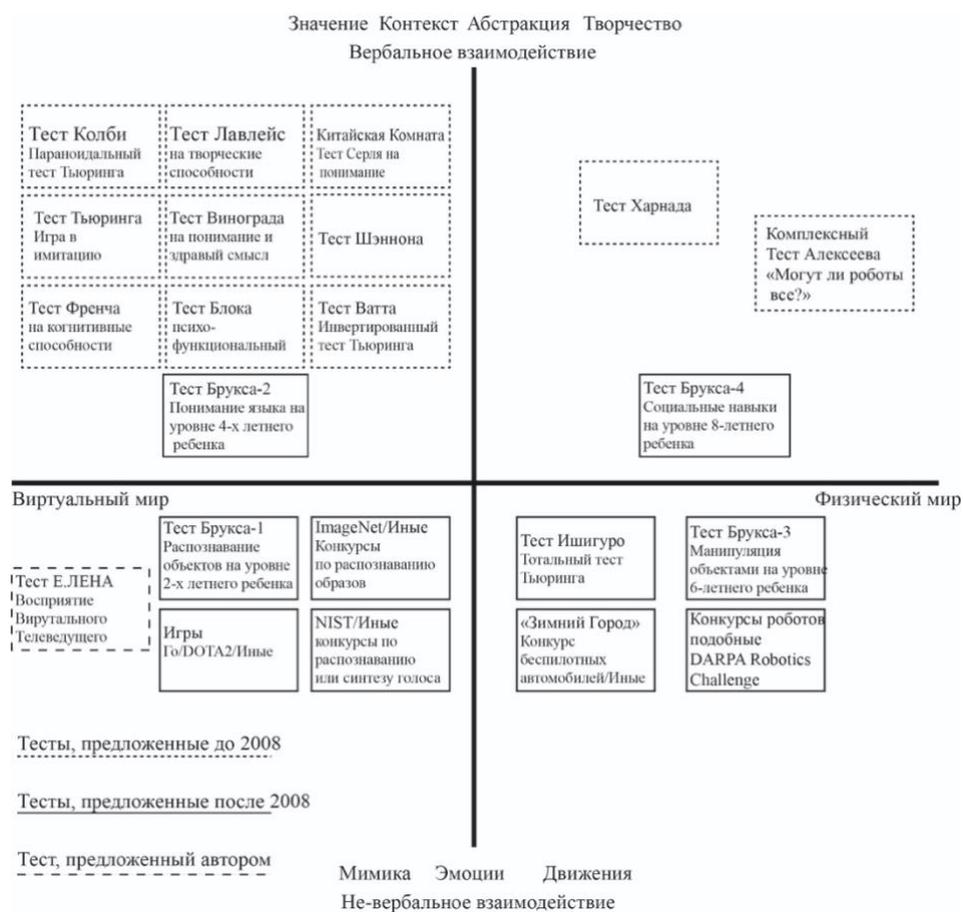


Рис. 1. Пространство тестов Тьюринга

Вербальное взаимодействие в виртуальном мире. Описанная в работе история становления исследований ИИ показывает, почему большинство тестов (мысленных экспериментов), разработанных до 2008 г., находится в этом квадранте. Проверка различных вербальных способностей при общении выступает основой в каноническом тесте Тьюринга, тесте леди Лавлейс, тесте Колби, тесте Сёрла, тесте Блока. Человек во всех этих тестах действует в виртуальном мире своего воображения и компьютерной программы. Интерфейс стандартный: экран, клавиатура, мышка. Пример аналогичного взаимодействия — общение пользователя с банковской программой. Казуальные связи в уже созданных системах понятий выступают главным основанием ответов машины, даже если компьютер в состоянии оценить социальные мотивы человека-собеседника.

Вербальное взаимодействие в физическом мире. Ещё А. Тьюринг отверг его использование, потому что оно очень ограничено использовалось

исследователями. Тесты, основанные на демонстрации вербального взаимодействия искусственного интеллекта и человека в физическом мире, предложили только С. Харнад и А. Ю. Алексеев. Сейчас распространена практика голосовой отдачи команд, например, звуковым колонкам, пылесосам и прочим устройствам, равно как аудиосообщений от контрольных терминалов или банкоматов людям.

Невербальное взаимодействие в виртуальном мире. Примером подобного взаимодействия выступают поединки между игровыми персонажами в компьютерных играх. Несмотря на то, что А. Тьюринг указал на его важность, эта область тестов долгое время не попадала в фокус внимания исследователей. Распознавание образов, человеческой речи, а также их синтез может быть примером невербального взаимодействия в виртуальном мире. Воздействие в физическом мире принципиально отсутствует, и даже если машина распознает речь человека (например, игрока), то она определяет сами лишь слова, но не их смысл. Примером невербального взаимодействия в виртуальном мире могут быть действия или эмоции виртуальных аватаров, которые несут большую смысловую нагрузку при полном отсутствии передачи вербальной информации.

Невербальное взаимодействие в физическом мире. В предельно упрощённом виде это автоматический шлагбаум, который должен подниматься, когда компьютер с помощью камеры узнаёт лицо человека. Всё резко усложняется, когда надо подражать действиям человека или робот должен свободно перемещаться по квартире или больничному коридору, передавать людям посылки, принимать от них предметы. Создаваемое людьми виртуальное пространство обладает исчисляемыми, программируемыми характеристиками, причём весьма ограниченными по своему разнообразию. На этом фоне реальность неисчерпаема. В физическом мире резко возрастает роль случайности, а проведение абстрагирования становится отдельной задачей. Этот квантиль области тьюрингоподобных тестов является самым сложным, ведь действие в нём напрямую зависит от сочетания технологий

искусственного интеллекта и робототехники. Исследователи уже с 40-х годов прошлого века просто игнорировали эту область, и пример подал сам А. Тьюринг. Между тем её значимость для коммуникации людей самоочевидна (например, жестикация) и подчеркивается всеми исследователями коммуникаций. Один из потенциальных тестов на уровень развития робототехники состоит в сравнении андроида и человека: машина произносит лишь предварительно записанные фразы человека, но обладает максимальным с ним сходством (Х. Исигуро).

Также для определения интеллектуальных способностей машины можно использовать тест Брукса, либо же тест Э.ЛЕНЫ — в их основе невербальное взаимодействие. И если тест Э.ЛЕНА находится в области, которая характеризуется невербальным взаимодействием в виртуальном мире, то тест Брукса, вероятно, максимально сложный в данном контексте, находится во всех четырёх областях пространства тьюрингоподобных тестов.

Теперь, когда мы определили и описали пространство тьюрингоподобных тестов, а также дали первичную классификацию (впервые) частных тестов Тьюринга, появляется возможность рассмотреть некоторые особенности этого пространства сквозь оптику существующих эпистемологических концепций. В частности, постнеклассическая рациональность позволяет нам увидеть «системные качества, не сводимые к свойствам элементов»¹⁶⁴ или отдельных областей. В. С. Степин отмечал, что в системах гомеостатического характера «возникают внутреннее пространство и время, отличные от внешнего физического пространства»¹⁶⁵ и подчеркивал, что это уже известно в биологии. Этот подход впервые в нашей литературе

¹⁶⁴ *Степин В. С.* Постнеклассическая рациональность и информационное общество // *Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта, (17-18 марта 2016 г., г. Москва, философский факультет МГУ им. М.В. Ломоносова).* М.: ИИнтелл, 2017. С. 64.

¹⁶⁵ Там же. С. 64.

перенесен на классификацию частных тестов Тьюринга в соответствии с более ранними идеями, основанными на идее *умвельта*.

Диссертантом предложено соотнести всё множество частных тестов Тьюринга с понятием *умвельт* (Я. Иксюль, Е. Н. Князева, В. Г. Буданов)^{166,167} для робота. В диссертации вводится понятие *техно-умвельт*, которое определяет мир робота в тьюринговом пространстве и содержит срез отображения мира, каким его «видит» робот, искусственный интеллект (или машина), а также набор действий, доступных для данного интеллектуального робота (или машины). Согласно предложенному определению, все частные тесты Тьюринга классифицируются по четырём видам техно-умвельтов, расположенным вдоль двух фундаментальных осей, уже описанных ранее: *виртуальное-физическое* и *невербальное-вербальное*. Таким образом, можно показать, что все созданные ранее или создаваемые в будущем частные тесты Тьюринга классифицируются в четыре техно-умвельта, образующих тьюринговое пространство человеко-машинного взаимодействия: **1) виртуальное-вербальное; 2) виртуальное-невербальное; 3) физическое-невербальное; 4) физическое-вербальное**. Каждое из этих пространств характеризуется собственным срезом восприятия и множеством доступных действий, которые проявляются в тестах Тьюринга для робота, способного к действиям в данном техно-умвельте. Ранее близкая конструкция умвельта была предложена В. Г. Будановым¹⁶⁸. В работе 2018 г. он предложил разделение умвельтов на четыре мира: *природный, сетевой, виртуальный, техно (машинный)*. Эти умвельты могут являться фундаментом для

¹⁶⁶ Термин заимствован из этологии (Иксюль).

¹⁶⁷ Буданов В. Г., Асеева И. А. Умвельт-анализ и дорожные карты Большого антропологического перехода // Материалы междунар. науч. конгресса «Глобалистика-2017». М.: ФГП глобальных процессов МГУ им. М.В. Ломоносова, 2017. URL: https://lomonosov-msu.ru/archive/Globalistics_2017/data/section_6_10143.htm (дата обращения: 20.06.2020).

¹⁶⁸ Буданов В. Г., Аршинов В. И., Лепский В. Е., Свирский Я. И. Сложность и проблема единства знания. Вып. 1: К стратегии познания сложности. М.: ИФ РАН, 2018. С. 9.

предложенных в настоящей работе техно-умвельтов, в которых реально работают роботы или машины, наделённые искусственным интеллектом.

Рассмотрим техно-умвельт, в котором действует компьютер (его обычно называют чат-бот) в классическом тесте Тьюринга. Данный техно-умвельт характеризуется а) используемым языком; б) набором понятий; и в) скоростью ответов (темпоральный аспект). В отличие от этого «*символьного*» техно-умвельта, который находится в виртуальном мире, техно-умвельт робота, который должен проходить сложные задания в физическом мире (DARPA Robotics Challenge, Зимний город) содержит множество способов отображения окружающего мира, при этом некоторые образы могут быть не видны для человека, так как робот использует весь электромагнитный спектр для построения навигационных карт. Однако робот, построенный для одного техно-умвельта способен действовать только лишь в этом техно-умвельте. Сравните с человеком, универсальность восприятия которого превышает искусственный интеллект любого робота, — мы способны двигаться (хотя и медленно) по любому типу поверхности, тогда как робот, созданный для офиса, по горам ходить не может. Робот, который воспринимает мир исключительно с помощью радиоволн, которые испускаются и поглощаются радарными, не может понять красный цвет. Но тем не менее, его техно-умвельт в радарном зрении имеет право на существование. С этим соглашаются и зарубежные авторы, дискутирующие вопросы техно-умвельта роботов и семиотического значения восприятия и искусственного интеллекта (см. W. Nöth¹⁶⁹ и C. Emmech¹⁷⁰).

Тем не менее, необходимо указать, что техно-умвельт робота не может однозначно соответствовать умвельту живого существа. Согласно Я. Иксклюю, умвельт животного содержит собственное целеполагание, «назначение жизни»¹⁷¹. Тогда как целеполагание робота, как выше указывал

¹⁶⁹ Nöth W. Semiosis and the Umwelt of a robot // Semiotica. 2001. No. 134. P. 696.

¹⁷⁰ Emmech C. Does a robot have an umwelt? // Semiotica. 2001. No. 134. P. 653-693.

¹⁷¹ Nöth W. Semiosis and the Umwelt of a robot // Semiotica. 2001. No. 134. P. 696.

В. К. Финн, не может содержаться в нём самом, а, следовательно, и в любом из умвельтов, которые строятся на основе его восприятия. Это качественное отличие, очевидное сейчас, не абсолютно, не задано раз и навсегда, потому что при создании технологических комплексов, способных к самовоспроизводству, будут получены роботы, обладающие внутренним целеполаганием.

2.3. Стена Тьюринга как фундаментальный философско-методологический барьер развития интеллектуальной робототехники

Ранее, в параграфе 2.2 среди элементов классического теста Тьюринга, предложенного в 1950 г. была указана «стена», которая разделяет участников (Ч и К) и судью (С), так как они должны находиться в разных комнатах. Сам Тьюринг указывал, что «новая постановка нашей проблемы имеет то преимущество, что позволяет провести чёткое разграничение между физическими и умственными возможностями человека»¹⁷². Однако преимущество было подвергнуто сомнению. В частности, С. Харнад считает, что это преимущество могло бы быть реализуемым, если бы разграничение касалось внешнего вида и действия компьютера. Но, так как линия установлена между вербальным и невербальным взаимодействием, то такое разграничение очень условно¹⁷³. По мнению С. Харнада, если бы у А. Тьюринга была возможность более тщательно продумать вопрос, то разграничение проходило бы между физическим обликом испытуемого и его способностями (как вербальными, так и невербальными). Если интеллектуальный робот или Общий искусственный интеллект должны чем-то отличаться от человека, то оценка различий должна быть на основе того, что они делают, а не как они выглядят.

¹⁷² Тьюринг А. Может ли машина мыслить. М.: ГИФМЛ, 1960. С. 13.

¹⁷³ Harnad S. Commentary on «Computing Machinery and Intelligence» // «Parsing The Turing Test» Philosophical and Methodological Issues in the Quest for the Thinking Computer / Ed. R. Epstein, G. Roberts, G. Beber. Amsterdam: Springer Netherlands, 2008. P. 27.

Несмотря на то, что буквально, терминологически «стена» не упоминается А. Тьюрингом напрямую, в своей статье он фактически описывает условие, которое требует от задающего вопросы — не видеть других участников игры, не слышать их голоса и физически не касаться их. Назовём это условие «стеной Тьюринга».

А. Тьюринг поставил другое ограничение на машины, которые принимают участие в игре в имитацию. Он считал, что только «цифровые компьютеры» могут принимать участие в этом тестировании¹⁷⁴. Однако, по нашему мнению, это произошло только лишь потому, что во время А. Тьюринга не было никаких других типов компьютеров, например, интеллектуальных роботов¹⁷⁵.

Стена Тьюринга выполняет следующие функции:

1. Разделяет субъекта испытания, судью и объекты испытания (человек или робот, компьютер).
2. Позволяет исследователям игнорировать все возможные техно-умельцы взаимодействия человека и робота, кроме одного — *виртуально-вербального*.
3. Выводит из области тьюринговой методологии трудные вопросы, такие как наличие сознания, невербальная коммуникация.
4. Противопоставляет деятельность компьютера (робота) и человека в их сравнении.

Последняя причина заставляет исследователей разрабатывать и реализовывать в тесте Тьюринга только те задачи, которые могут быть решены роботом в полном отрыве от человека. Ситуация во многом сходна с отношением, которое было у общества к машинам в эпоху первой промышленной революции. Согласно мифологической истории, молотобоец

¹⁷⁴ Turing A. Computing machinery and intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 554.

¹⁷⁵ Harnad S. Minds, Machines and Turing: The Indistinguishability of Indistinguishables // Journal of Logic, Language and Information. 2000. Vol. 9. P. 425–445.

Джон Генри победил в соревновании с паровой машиной, но умер от истощения¹⁷⁶.

Стена Тьюринга делает легко представимой ситуацию, в которой компьютер может превзойти человека в тестируемой способности. Это может быть вербальное общение или зимнее вождение автомобиля. Смысл работы исследователей и разработчиков становится в том, чтобы добиться степени превосходства системы, в которой тестируемый и тестирующий разделены стеной.

Однако современные исследования в различных областях человеко-машинного взаимодействия (В. Унхелькар, Дж. Ша и др.)¹⁷⁷ говорят о том, что максимальный эффект от использования роботов или искусственного интеллекта достигается при совместном использовании, т.е. в коллаборации человека и машины. Исследования показывают, что связка робот и человек может иметь большую эффективность при выполнении практической задачи. Иными словами, достижение целевой ситуации классического теста Тьюринга — компьютер способен поддерживать диалог с человеком — может быть менее эффективно, чем достижение общей цели в совместном диалоге человека и машины. Исследования коллаборации человека и интеллектуального робота показывают, что в типовых случаях время ожидания действия человека снижается на 41%, а время совместной работы робота и человека возрастает на 71% по сравнению с теми же операциями, в которых человек и робот действуют по отдельности, без координации друг с другом¹⁷⁸.

¹⁷⁶ Джон Генри // Сказки народов мира. URL: <http://skazki-narodov.ru/181-dzhon-genri.html> (дата обращения: 16.08.2020).

¹⁷⁷ *Unhelkar V. V., Shah J. Enabling Effective Information Sharing in Human-Robot Teams // Conference Robotics: Science and Systems (RSS), 2018.* URL: http://interactive.mit.edu/sites/default/files/documents/Unhelkar_RSS_Pioneers_Abstract_2018.pdf (дата обращения: 16.07.2020).

¹⁷⁸ *Nikolaidis S., Lasota P. A., Rossano G., Martinez C., Fuhlbrigge T., Shah J. Human-Robot Collaboration in Manufacturing: Quantitative Evaluation of Predictable, Convergent Joint Action*

Инженерным коллективом под руководством диссертанта был разработан экспериментальный стенд по отработке сценариев коллаборативного взаимодействия робота и человека и получен патент на уникальное решение, позволяющие роботу и человеку безопасно работать в едином пространстве¹⁷⁹.

Стена Тьюринга заставляет исследователей создавать решения, которые ориентированы на работу без участия человека, тогда как участие человека (в любом виде деятельности) может оказать кардинальное влияние на работу общей связки человек-робот/компьютер.

Разделение человека и машины при операционалистическом сравнении является ненужным ограничением, которое тормозит развитие всей области искусственного интеллекта и технического прогресса. Необходимо сравнение результативности действий совместной работы «связки» робота (в общем случае искусственного интеллекта) и человека, в отличие от результативности действий человека самого по себе. Жёсткое взаимоотношение и взаимоисключение порождает неразрешимый конфликт, который препятствует решению задач создания Общего искусственного интеллекта. Но этот конфликт по своей сути и сама стена Тьюринга связаны как раз именно с игнорированием проблемы сознания.

Асимметричность взаимодействия человека и машины задаёт другую крайность, а именно чрезмерное, контрпродуктивное разделение функций. Человек начинает полагаться на ответ компьютера буквально во всём и утрачивает: а) собственную целостную картину мира, б) собственное целеполагание. Этот эффект наблюдается уже сегодня, при чисто символическом человеко-машинном общении: студенты утрачивают навыки запоминания

// International Symposium on Robotics (ISR), 10/2013. URL: <http://interactive.mit.edu/human-robot-collaboration-manufacturing-quantitative-evaluation-predictable-convergent-joint-action> (дата обращения: 10.07.2020).

¹⁷⁹ Патент недели: может ли робот причинить вред человеку // Популярная механика. URL: <https://www.popmech.ru/technologies/487412-patent-nedeli-mozhet-li-robot-prichinit-vred-cheloveku/> (дата обращения: 10.07.2020).

информации, формулировки ответов при очевидных посылках, как следствие, в пределе, они утратят навыки формулировки проблем и будут вбивать в поисковик лишь собственные неудобства («жарко, что делать?»). В результате человек утрачивает полноту процесса абстрагирования, подстраивается под символическую конструкцию и набор алгоритмов поведения, которые предлагает машина. Это вызывает закономерные опасения, что создание «сильного ИИ» приведёт либо к созданию мрачных политических диктатур, либо вообще, к утрате человеком, как видом, главенствующих позиций в техносфере.

Кроме того, в символическом потоке данных, который получает человек, проще лгать, замалчивать сведения. Естественно, компьютер может не только умышленно лгать, но и неадекватно воспринимать действительность: А. М. Анисов, анализируя генетический метод построения теорий с использованием машины Тьюринга, показывает, что «вычисление, рассуждение и доказательство», как различные понятия, порой не различаются исследователями, а «само по себе построение абстрактных объектов, если о них ничего доказательно не утверждается, бессмысленно»¹⁸⁰. Также символическое общение сталкивается с проблемами передачи личностного знания: «Художник может заметить такие цветовые оттенки на стволе и ветвях, которые незаметны для всякого другого. Передать в словах индивидуальные особенности восприятия данного дерева, его «личный образ» довольно сложно. Ведь в языке выражается только общее, коллективное»¹⁸¹.

Из сказанного выше становится ясно, что общение человека и машины может происходить не только в форме чисто абстрактных, печатных символов, оно уже сегодня сдвигается к физическому и невербальному взаимодействию.

Символы вербализуются — машина может озвучить текст, равно как и превратить речь пользователя в письменные строчки. Это уже предполагает

¹⁸⁰ Анисов А. М. Генетический метод построения теорий // *Credo new*. 2012. № 3. С. 43.

¹⁸¹ Лекторский В. А. Эпистемология классическая и неклассическая. М.: Эдиториал УРСС, 2001. С. 194.

снижение, размытие «стены Тьюринга», потому что текст, озвученный сам по себе, без эмоциональных коннотаций, воспринимается сложно. И программы-чтецы, в итоге, должны соотносить произношение и содержание текста. Но невербальное взаимодействие, пусть только с помощью звука, уже может не сводиться к символам, как таковым. Машина может громким зуммером в критическую минуту оповестить пользователя о тревоге не столько в символической форме, сколько в эмоциональной. Может передавать длительный поток данных, направленный исключительно на эмоциональное воздействие, например мурчание кошки или изображение красивых пейзажей, чтобы успокоить пользователя. Подобные действия — начало не только устранения стены Тьюринга, но и формирования трансдисциплинарной эпистемологической позиции машин, которая позволит в будущем достигнуть «единства знаний»¹⁸², снять для техники барьеры между научным и культурным познанием.

Но это происходит, если машина начинает воздействовать на эмоции пользователя. Простые алгоритмы, внешние по отношению к индивиду, сводят восприятия человека к чисто бихевиористским стандартам. Если речь идёт о двустороннем осознанном взаимодействии с психикой, то машине необходимо оценивать эмоции. Да, частично такое уже воплощено в системах распознавания лиц, и будет развиваться в ближайшее время, когда данные, взятые с фитнес-браслета (пульс, уровень кислорода и т.п.) будут расшифровываться, как определённая эмоция пользователя, и запускать контекстную рекламу на мониторе компьютера. Но направленность эмоций, их связь с объектами невозможно понять только в мгновенном снятии данных, и самые совершенные алгоритмы распознавания мимики здесь не помогут, потому что они в принципе не могут учесть всё разнообразие окружающих человека объектов. Задача эта будет решаться через соотношение «историческое-логическое», то есть машина должна знать историю развития

¹⁸² *Киященко Л. П.* Личность как голограмма в трансдисциплинарной культуре // Вопросы философии. 2017. № 11. С. 58–68.

чувств человека, увязывать те или иные проявления эмоций с общим социально-культурным контекстом, образом жизни, биографией и т.п. Примером тут может быть шахматное соревнование человеко-машинных гибридов: компьютер будет оценивать позицию на доске, одновременно анализировать проявление эмоций оппонента, в то время как человек, при условии относительно равных возможностей электронных устройств обоих игроков, будет искать нетривиальные линии поведения, неожиданные ходы. Г. Каспаров называет таких игроков «кентаврами, командами, которые являются сплавом людей и алгоритмов»¹⁸³.

Тут возникает вопрос: не будет ли человек еще более уязвим по отношению к машине, если будет получать заведомо структурированную информацию, вербальную и невербальную. Воздействие на тело тоже будет структурировано в рамках общения с машиной. И весь этот поток сигналов будет контекстно увязан с его биографией? Почему не смогут (или не должны) воплотиться образы фантоматической машины С. Лема, фильма «Матрица» и прочие дистопические картины?

Сейчас пользователи получают избыточный поток данных, поскольку они как личности, как отдельные субъекты, взаимодействуют со стремящимися к бесконечности электронными библиотеками, с коллективными субъектами больших фирм и государства. Телефон, который наговаривает обывателю печатные тексты, в субъектном отношении является скорее маркетинговым инструментом по управлению человеком, чем инструментом самого пользователя.

Показателем субъектности пользователя становится та доля в цепочке превращения феномена в абстракцию, в движении маятника «идеальное-реальное», которую контролирует именно человек. Сочетание целостности (возможности её самоподдержания), которой мог бы обладать робот, и

¹⁸³ The dawn of artificial intelligence // The Economist. Weekly edition. May 9, 2015. URL: <https://www.economist.com/leaders/2015/05/09/the-dawn-of-artificial-intelligence> (дата обращения: 19.07.2020).

заведомой ограниченности в доле этой цепочки дают образы животного. Лучший из существующих помощников человека — это собака: стайное животное, частично социализированное, которое в большей степени получает свою целереализацию благодаря человеку. Возникает вопрос: если компьютерная революция приводит к такой алгоритмизации человеческой жизнедеятельности, сумма которой обеспечивает правила абстрагирования, формулирует цели развития личности и социума, то как тогда человеку, индивиду, сохранить позиции в контроле маятника «идеальное-реальное»?

Прямое копирование образа животных при конструировании роботов не имеет смысла, так как животные ограничены в усвоении абстракций, а компьютеры изначально создавались именно для работы с абстракциями. При относительно мобильном теле и неограниченной, безлимитной беспроводной связи вы получаете не субъекта-животное, киборга-собаку, а робота-аватара, которым управляют малопонятные алгоритмы. Но сейчас малореально даже создание относительно гибкого, с большим количеством степеней свободы робота, хотя бы на уровне двухлетнего ребёнка. И тут мы переходим к проекции «стены Тьюринга» на физическую активность машин — к «парадоксу Моравека».

Сформулированный во время «зимы ИИ 80-х годов» — периода сниженных ожиданий в развитии ИИ — парадокс Моравека, как нельзя лучше иллюстрирует ограниченность машин: «относительно легко достичь уровня взрослого человека в таких задачах как тест на интеллект или при игре в шашки, однако сложно или невозможно достичь навыков годовалого ребёнка в задачах восприятия или мобильности»¹⁸⁴. Это ограничение, по сути, в конкретизации действия любой алгоритмической машины: любая не структурированная, слабо описанная ситуация становится препятствием для исполнения алгоритма.

¹⁸⁴ *Moravec H. Mind Children: The Future of Robot and Human Intelligence. Cambridge, MA: Harvard University Press, 1988. P. 15.*

Если предположить, что ни стена Тьюринга, ни парадокс Моравека не будут преодолены (хотя уже сейчас имеются существенные аргументы для их опровержения), то следует допустить сценарий создания машин, которые превосходят человека в работе с символьными переменными, или, в общем случае, с любыми репрезентациями или абстракциями — именно человеку, индивиду неизбежно останется роль, которой так боялся Л. Мэмфорд — «младшего партнёра», придатка к машине. Человек будет обречён на то, что ему достанется вся грязная, скучная работа. Из программиста-оператора, который во времена формулировки теста Тьюринга мог сам исправлять программный код в случае замеченных ошибок, человек скатывается к роли ремонтника и отчасти тестировщика, который должен подтверждать или опровергать решения машин в неcodифицированных ситуациях. На сегодняшнем уровне развития техники программа, которая управляет беспилотным автомобилем, не может играть в шахматы, а человек, который, казалось бы, всегда общается с частной, узкопрофильной программой, сохраняет свободу воли. Но совокупность этих программ и алгоритмов, которую проще всего проиллюстрировать впечатляющим количеством законодательных запрещений и необходимых для любой деятельности документов — будет ограничивать человека в продуктивной работе с абстракциями. Функционализм, как критерий успешности в деятельности «слабых ИИ», при этом рассматривается как целиком действующая доктрина.

Как тогда сохранить баланс в человеко-машинном взаимодействии?

Если интеллектуальные роботы или компьютеры в когнитивном смысле сотрудничают со своими пользователями не только в подтверждении слабо алгоритмизированных ситуаций, а им приходится использовать абстракции, созданные данным конкретным человеком, то становится невозможна полная централизация работы с абстрактными понятиями, с цифрами. То есть потенциально достижима локальная субъектность пары, в которой машина будет взаимодействовать с человеком. Необходима *локализация субъектности* компьютера, даже если набор процессоров и ячеек памяти,

который управляет машиной, находится «в облаке». И таким путём может осуществляться физическое, невербальное, эмоциональное взаимодействие компьютера с человеком. Возможно, это потребует дополнительных мер по защите личной информации, но в сложившейся паре и робот, и человек смогут одновременно работать на разных уровнях процесса «абстрагирования-конкретизации».

Кроме того, возможно взаимодействие человека и машины по принципу двойного субъекта В. Е. Лепского: машина фактически предоставит человеку среду обитания (накопления знаний), в которой будет идентифицирован виртуальный субъект¹⁸⁵.

Сложно сказать, насколько глубоким может быть взаимопроникновение мышления человека и машины в бесшовной коллаборации, но непосредственный «напарник» человека будет так же заинтересован в качестве работы. Образно, программы для генерации текстов дают возможность пользователю больше внимания уделять концептуальному подходу, выстраиванию общей структуры произведения. Сейчас это слишком часто приводит к простому копированию, либо рерайтингу чужих текстов. Но представим, что машина воспринимает эмоциональное переживание текста своим пользователем, может невербально общаться с ним в процессе создания текста. Тогда целью может стать не максимально быстрое копирование, а последовательное конструирование идеи, создание структуры произведения — но это возможно, если произведение отвечает истории общения пользователя с другими текстами, его общим навыкам и склонностям. То есть пара машина-человек, чья связность будет выше, чем у служебной программы на этом компьютере и сервисного центра по её обслуживанию, — эта пара будет стремиться выполнять развивающую именно её работу.

Еще одним интересным аспектом проявления стены Тьюринга в общении человека и компьютера является снижение значимости барьера

¹⁸⁵ Лепский В. Е. Аналитика сборки субъектов развития. М.: Когито-Центр, 2016. С. 50.

«зловещей долины», о которой говорилось выше. В случае если компьютер (или интеллектуальный робот) приобретает некоторые свойства человека по передаче вербально-невербальной информации, то возникает эмоциональная привязанность человека к компьютеру-роботу-программе, о которой писал М. Маклюэн еще на заре компьютерной эры в 1964 г.: отражение Нарцисса — это же его расширение, удвоение, в которое влюбился юноша; так и техника постоянно модифицирует человека, выполняя его желания, а люди превращаются в органы размножения машинного мира, влюбляясь в технику, выполняющую их желания¹⁸⁶.

М. Маклюэн предугадал появление эмоциональной привязанности человека не только к гаджетам, как таковым, но и к той особенной технологической сущности, которую сейчас принято называть «голосовыми виртуальными ассистентами». Данные устройства сочетают несколько видов распознавания жестов, действий и голоса, и, за счёт комплексирования этих данных, способны к распознаванию эмоций человека и настройке своего диалога в соответствии с настроением пользователей. Такие устройства обладают возможностью эмоциональной привязки пользователя к своим сервисам, подобно тому, как Нарцисс был привязан к своему собственному отражению.

Предполагается, что стена Тьюринга, которая отделяет тестируемый компьютер (или виртуальный ассистент) теряет релевантность в данном случае, так как функциональная полезность гаджета для пользователя не меняется в зависимости от того, насколько точно он имитирует действия человека (коммуникации). Скорее наоборот, привязанность к виртуальному ассистенту возрастает в зависимости от его полезности для пользователя, а не его похожести на человека, подобно тому, как привязанность Нарцисса к собственному отражению никак не связана с тем, что он осознает, что смотрит

¹⁸⁶ Маклюэн М. Понимание Медиа: внешние расширения человека. М.: Гиперборей, 2007. С. 112–113.

на свое отражение — мы знаем, что это лишь отражение, но тем не менее все равно привязаны к нему.

И тут потенциальное противоречие между машиной-помощником и человеком может быть снято не только если машина приобретет неотличимую от человека антропоморфность, но ещё если коллективная субъектность в паре человека-машины (стремящаяся к индивидуальной, к слиянию) может быть сбалансирована за счёт использования машиной ощущений самого человека. Человек становится сенсором, «чувствилищем», некоей совокупностью органов чувств для машины, равно как машина предоставляет человеку поток сенсорной информации. Однако уже сейчас фотоаппараты проводят съёмку в инфракрасных лучах. Машины готовы пользоваться более широким спектром физических феноменов для оценки своего окружения и там, казалось бы, «человек-чувствилище» становится как человек-компьютер (т.е. расчётчик) 1940-х, который вынужден был менять профессию после появления машин-компьютеров. Зачем машинам поток ощущений от человеческих органов, если техника может передать видеопоток с высочайшим разрешением, фиксировать недоступные людям звуки?

Да, это так. Только в эмоциональных взаимосвязях пары человек-машина, индивид не проводит процесс абстрагирования до уровня понятий, и не выступает «оператором», вопрос именно в замыкании контура обратной связи не только со стороны человека, но и со стороны машины. Компьютер должен оценивать, насколько хорошо-плохо-вдохновенно-безрадостно его напарнику.

Субъектность человеко-машинной пары при этом служит сдерживающим фактором: «умный дом» может обходиться без человека-жильца. Но требуется не создать систему получения информации, заведомо лучшую, чем тело человека, а использовать сигналы человеческой нервной системы для длительной оценки информации, обработка которой в абстрактных понятиях пока невозможна или малопродуктивна.

Вышесказанное позволяет увидеть триаду постнеклассической рациональности субъект-средство наблюдения-объект в контексте роли стены Тьюринга, как эпистемологической преграды на пути взаимодействия человека-субъекта и робота-объекта. Ранее мы показали, что стена Тьюринга оказывает неизбежное влияние на весь цикл взаимодействия машины и человека, сдвигая фокус внимания на связку средства вопрошания (тест Тьюринга) и самой машины. Переход через стену Тьюринга в коммуникациях между роботом и человеком создает человекомерную систему, в которой робот и человек являются полноправными участниками, взаимодействующими друг с другом и познающими мир совместно, используя сознание человека-субъекта в реальном времени. Объектом исследования может быть не только сам робот, который проходит тест Тьюринга, но и внутренний мир человека, так как «тёмная материя»¹⁸⁷ культуры будет во многом определять взаимодействие с помощью невербального канала коммуникации. Такой подход восстанавливает расчленённое А. Тьюрингом с помощью стены взаимодействие робота и человека.

Подведём итоги. Преодоление стены Тьюринга подразумевает надделение машины значимым количеством человеческих функциональных особенностей, которыми программы и антропоморфные роботы сейчас не обладают. При этом открываются как возможности преодоления тупиков развития индивида, так и углубления кризиса современных форм взаимодействия человека с электронными устройствами. Потому остро стоит задача создания такой посттьюринговой методологии, которая бы позволила, с одной стороны, успешно создать Общий искусственный интеллект и интеллектуальных роботов, полезных, эффективно взаимодействующих с человеком, а с другой — избегать нежелательных последствий.

¹⁸⁷ Эверетт Д. Как начинался язык. История величайшего изобретения. М.: Альпина Нон Фикшн, 2019. С. 384.

Глава 3. Перспективы применения посттьюринговой методологии в интеллектуальной робототехнике

3.1. Посттьюринговая методология как концептуальный инструментарий разрушения стены Тьюринга

Ранее мы ввели и обсудили понятия посттьюринговой методологии, техно-умвельтов для искусственного интеллекта и интеллектуальных роботов, стены Тьюринга, а также общего пространства тьюрингоподобных тестов. Ниже мы рассмотрим связь этих понятий и их возможное использование для разрушения стены Тьюринга.

Как ранее отмечалось в Главе 1, интеллектуальная робототехника играет ту же концептуальную роль для Общего искусственного интеллекта, что человеческое тело играет для естественного интеллекта. Роль воплощённого сознания и воплощённого интеллекта отмечалась еще Дж. Локком, который писал, что «структуры, образующие нашу концептуальную систему, имеют своим источником наш чувственный опыт и осмысляются в его терминах»¹⁸⁸. Первые эмпирические опыты по исследованию интеллекта, воплощённого в работе, начал проводить американо-британский нейрофизиолог и кибернетик, сооснователь Клуба Рацио (совместно с А. Тьюрингом), У. Грей в 40–50-х годах прошлого века. Он разработал электромеханических черепах, которые, обладая эквивалентном лишь двух нейронов, могли полностью автономно передвигаться, реагируя на свет. Черепахи пользовались большой популярностью и известно, что А. Тьюринг внимательно изучал этих примитивных роботов и дал восторженный отзыв в 1951 г.¹⁸⁹

¹⁸⁸ *Лакофф Д.* Женщины, огонь и опасные вещи. Что категории языка говорят нам о мышлении. М.: Языки славянской культуры, 2004. С. 13.

¹⁸⁹ *Boden M.* Grey Walter's Anticipatory Tortoises // The Rutherford journal. URL: <http://rutherfordjournal.org/article020101.html#null> (дата обращения: 16.06.2020).

Аналогичная работа по изучению воплощенного интеллекта велась параллельно в СССР в Институте машиноведения РАН и Институте автоматизации и телемеханики¹⁹⁰.

Важный шаг в развитии теории «воплощённого интеллекта» был сделан французским философом М. Мерло-Понти, который предположил, что интенциональные действия человека не обязательно должны сопровождаться символической репрезентацией, но могут быть выучены, натренированы и воплощены самим организмом¹⁹¹. Концепция воплощённого искусственного интеллекта находит своё отражение во многих работах современных исследователей как в области искусственного интеллекта, так и воплощённой, ситуативной робототехники (Р. Брукс, Б. Херцль и др.). Однако ранее все предложения исследователей, создающих роботов, основанных на «воплощённом искусственном интеллекте», концентрировались лишь на одном избранном умвельте, который зачастую отражал лишь специализацию авторов той или иной концепции. К примеру, Х. Ишигуро, известный японский исследователь робототехники и человеко-машинного взаимодействия, считает, что достаточно лишь «имитировать» физический облик человека. При этом смысл речи, образов и действий партнёров-людей остаются вне техно-умвельта роботов, которые создает этот исследователь. По сути, роботы Х. Ишигуро находятся в техно-умвельте *физическое-невербальное*, который не подразумевает наличия в нём людей, желающих взаимодействовать с роботом на уровне смысловых понятий. Для исследователя важно лишь то, что восприятие людей «обмануто» иллюзией общения. Похожая ситуация возникает и с использованием мощных возможностей искусственного интеллекта по обработке текстовой информации. Искусственные нейронные сети, которые обладают числом синапсов, приближающимся к человеческому мозгу (175 млрд) и демонстрирующих прекрасные имитации текстов, не в состоянии выполнить

¹⁹⁰ Кобринский Н., Пекелис В. Быстрее Мысли. М.: Молодая Гвардия, 1963. 443 с.

¹⁹¹ Merleau-Ponty M. Phenomenology of perception. London: Kegan Paul, 1962. 544 p.

задачи, которые решали черепахи У. Грея более 70 лет назад с помощью лишь двух нейронов.

Простыми примерами воплощённой робототехники являются интеллектуальные роботы, наподобие черепах У. Грея или современных роботов-пылесосов, созданных на основе ситуативного подхода Р. Брукса¹⁹². Они управляются простыми нейронными сетями, например, однослойной нейронной сетью без какой-либо внутренней организации. Тем не менее, эти роботы могут демонстрировать не только простое поведение, такое как избегание препятствий, но и поведение, способное решать сложные проблемы, включая избегание искажения восприятия, разрешение сенсорной неоднозначности и последовательную организацию подчиненных поведений. Возникает вопрос, насколько далеко мы можем зайти с реактивной сенсомоторной координацией, анализом ситуации и общими ограничениями подобного подхода.

Разработка роботов или систем искусственного интеллекта лишь в одном из техно-умweltов приводит к четырём фундаментальным ограничениям в развитии как Общего искусственного интеллекта, так и интеллектуальных роботов.

Во-первых, ограниченность познавательной, когнитивной функции доступным пространством техно-умweltа — архитектура информационных систем, сенсоров и актуаторов робота не позволяет ему переносить приобретаемые знания, навыки на действия в других техно-умweltах. К примеру, пройденный тест Тьюринга на основе диалогов или выигранная партия у человека в настольную игру не позволяет роботу уверенно действовать в физическом мире. Стена Тьюринга, которая отделяет один техно-умwelt от другого через методологические ограничения, делает любой познавательный опыт искусственного интеллекта ограниченным без необходимости.

¹⁹² *Brooks R. A.* For the Last-Minute Holiday Shopper: Get an iRobot! URL: <https://corporature.wordpress.com/tag/rodney-brooks/> (дата обращения: 20.08.2020).

Во-вторых, в тьюринговой методологии человек имеет роль наблюдателя-судьи для действий робота или искусственного интеллекта в каждом отдельно взятом техно-умвельте. Однако, как было показано в параграфе 2.3 Главы 2, это не оправданно, так как действия связки «робот и человек» являются более эффективными, чем действия отдельно робота и человека. Стена Тьюринга, которая отделяет наблюдателя-судью-человека от робота, выполняющего определённые действия, делает неэффективными обоих — робота и человека.

В-третьих, роботы, созданные для разных техно-умвельтов, не в состоянии находить общий язык и общий техно-умвельт без переводчика — человека-инженера, который вынужден создавать программно-аппаратные мосты, позволяющие роботам выполнять простейшие совместные действия. Стена Тьюринга, заставляющая исследователей и инженеров создавать роботов лишь для одного техно-умвельта, не позволяет налаживать автономного взаимодействия роботов или систем общего искусственного интеллекта между собой.

В-четвертых, надёжность работы искусственного интеллекта или интеллектуальных роботов определяется исключительно теми ситуациями, которые возможны в рамках того техно-умвельта, для которого они созданы. Но мир, в котором действуют люди и создаваемая ими техника, охватывает все четыре техно-умвельта. Любое действие из одного техно-умвельта, которое может коснуться робота, который выполняет работу в другом техно-умвельте, приводит к неизбежному сбою. Стена Тьюринга выполняет роль имманентного препятствия, на которое любой робот будет наталкиваться в попытке решить задачу, пришедшую из другого техно-умвельта. К примеру, программа искусственного интеллекта, побеждающая в игру Dota2, никогда не сможет победить команду людей в пейнтбол без фундаментальной модификации. Аналогичный аргумент предлагал ранее Дж. Сёрл, когда

говорил, что «программа моделирования лактации и фотосинтеза не в состоянии дать нам молоко и сахар»¹⁹³.

Глубокой проблемой междисциплинарного значения становится вопрос, как может появиться и развиваться интеллект как новое свойство в группах роботов, аналогично тому, как интеллект появляется в группах социальных животных, действующих в одном биологическом умвельте. Согласно нашему определению интеллекта, сообщества роботов обладают интеллектом, если они могут решать проблемы определённой степени сложности, действуя в своем техно-умвельте. Вопрос может быть решён, если сосредоточить внимание на самых ранних этапах появления и эволюции простых технических артефактов. Следовательно, надо начать с построения искусственного общества воплощённых агентов в виде интеллектуальных роботов, создания среды или искусственной экосистемы и соответствующих примитивных форм поведения для этих роботов, а затем беспрепятственного запуска искусственного общества. Даже с небольшими популяциями простых роботов может быть произведено большое количество взаимодействий между роботами¹⁹⁴. Неоднородность, присущая интеллектуальным роботам, действующим в реальном техно-умвельте *физическое-вербальное*, а также шум и неопределённость реального мира увеличивают пространство возможностей для неожиданных проявлений во взаимодействии между роботами.

Одним из возможных способов решения указанных выше проблем является использование посттьюринговой методологии, которая предполагает: 1) прямое взаимодействие интеллектуальной компьютерной системы (*виртуально-вербальный* техно-умвельт) с человеком; 2) непрерывные переходы интеллектуальных роботов из одного техно-умвельта

¹⁹³ Сёрл Дж. Разумы, мозги, программы // Тест Тьюринга. Роботы. Зомби / Пер. с англ. Д. Родионова; под ред. А.Ю. Алексеева. М.: МИЭМ, 2006. С. 27.

¹⁹⁴ Brooks R. A. Cambrian intelligence: The early history of the new AI. Cambridge, MA: The MIT Press, 1999.

в другой. Для этого необходимо, чтобы создание когнитивных архитектур и их воплощение в механизмах, которые ориентированы на деятельность сразу в двух или больше техно-умвельтах — роботы или, в общем случае, искусственный интеллект — могли решать задачи в разных техно-умвельтах, исходя из единых инструментов, механизмов и архитектур. Конструирование конкретных систем должно включать в себя возможность последовательных действий в разных техно-умвельтах, оставаясь в рамках одной архитектуры и одного робота (или системы искусственного интеллекта). Данный подход значительно увеличивает сложность задачи в силу избыточности архитектуры и технических характеристик роботов вследствие универсализма разрабатываемых систем. Однако, именно такой подход может стать основой для прорыва в развитии Общего искусственного интеллекта и интеллектуальной робототехники, как бы повторяя биологическую эволюцию в отношении технических систем.

Представляется допустимой параллель посттыюринговой архитектуры интеллектуального робота с биологической эволюцией, в которой среда (биоумвельт) играла ключевую роль в адаптации биологических видов. Попробуем уподобить один из техно-умвельтов стихии, в которой возникла жизнь на Земле — Мировому океану. В этом случае, выход интеллектуальных роботов из первого техно-умвельта способен дать им новые, адаптационные свойства подобно тому, как «слепой часовщик» эволюции в течение миллионов лет давал новые возможности живым существам, вышедшим на сушу из океана. Переход к очередному техно-умвельту для роботов может означать следующий виток появления свойств. Речь идет не о том, что робот, который должен автономно ездить по суше, вдруг научится автономно плавать. Но речь идет о том, что способности робота, добившегося успеха в одном из техно-умвельтов, должны постепенно переноситься, как свойства в работе, в другой техно-умвельт. При этом в эволюционном цикле развития интеллектуальных, воплощённых роботов, усиливается роль создателя-

человека, который, наблюдая за ходом эволюции, может наделять роботов дополнительными техническими возможностями.

С. Лем в своей философской работе «Сумма технологий»¹⁹⁵ показал, что техника, в отличие от жизни, позволяет накапливать информацию без привязки к генетическому коду живущего организма, потому можно быстро развивать удачные решения. Сегодняшние технологии позволяют широко моделировать виртуальные тела для роботов. В частности, вопросы манипуляции предметами физического мира роботами наиболее эффективно решаются с помощью одного из видов машинного обучения, т. н. обучения с подкреплением. Обучение роботов новым навыкам выполняется с помощью создания симуляций физического мира и многократного повторения действий в этих симуляторах пока не достигается желаемый результат. Такой вид обучения стал возможным благодаря значительному удешевлению вычислительных мощностей и доступностью цифровых симуляторов¹⁹⁶. Кроме того, большой объем информации поступает от работающих машин (становление интернета вещей) и от людей, которые с ними взаимодействуют. Вероятно, в ближайшей исторической перспективе будет возможно моделирование эволюционных процессов, основанное на виртуальных моделях, единичных (контрольных) экземплярах роботов и на потоке данных от реальных технологических систем и социумов. Это позволит сравнительно быстро проверять выдвигаемые гипотезы и мультиплицировать удачные решения. И хотя скорость получения результатов неизбежно окажется ниже, чем у работы в рамках единственного техно-умвельта (например, создание новой криптографической программы), технологический субстрат для эволюционного развития существует.

¹⁹⁵ Лем С. Сумма технологии = Summa Technologiae: пер. с польск. / Пред. В.В. Ларина; ред. и послесл. Б.В. Бирюкова и Ф.В. Широкова. М.: Мир, 1968. С. 516.

¹⁹⁶ Лаборатория робототехники Сбербанка. Аналитический обзор мирового рынка робототехники. URL: https://www.sberbank.ru/common/img/uploaded/pdf/sberbank_robotics_review_2019_17.07.2019_m.pdf (дата обращения: 21.07.2020).

Концептуально архитектура постъюрингового интеллектуального робота в схематическом виде представлена на рис. 2

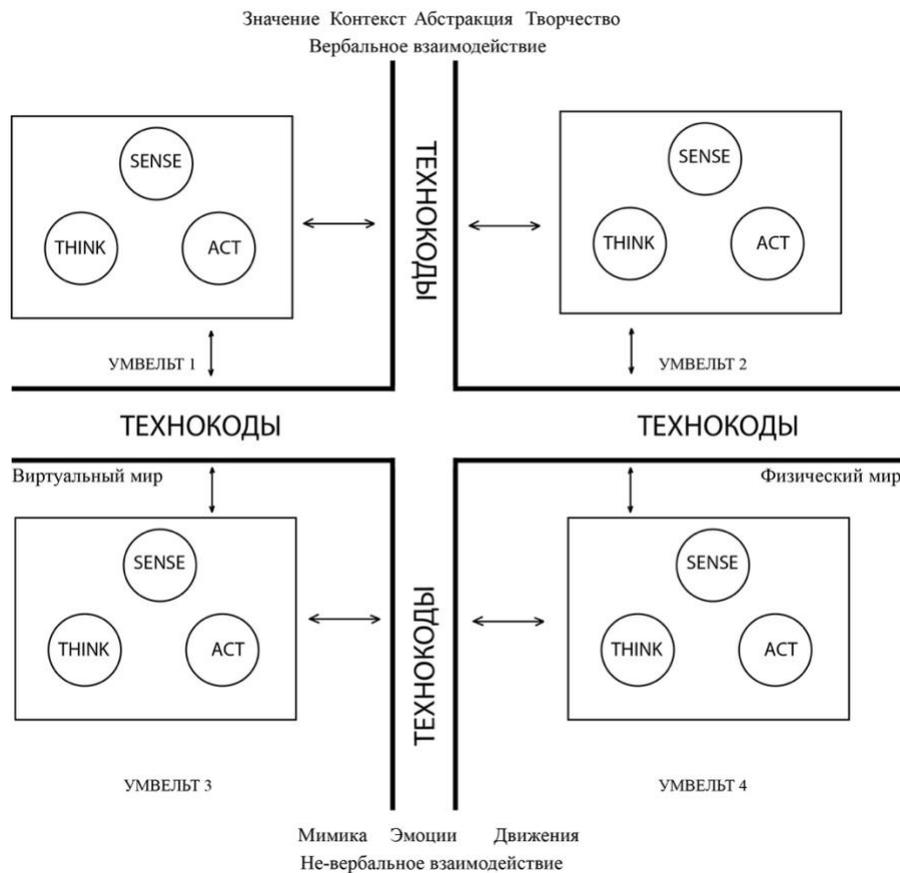


Рис. 2. Концептуальная архитектура постъюрингового интеллектуального робота

Раскроем содержание основных элементов такой архитектуры:

1. Робот обладает способностями к ощущениям, пониманию и действиям в каждом из техно-умвельтов: 1) *виртуальное-вербальное*, 2) *физическое-вербальное*, 3) *виртуальное-невербальное*, 4) *физическое-невербальное*. Для краткости они обозначены как 1, 2, 3 и 4.
2. Робот обладает независимой возможностью совершать осмысленные действия в каждом из техно-умвельтов с помощью специализированной архитектуры, которая эффективно работает в данном техно-умвельте. Данную архитектуру назовём умвельт-машиной. Робот может иметь в своем составе одну или несколько умвельт-машин (до четырёх).

3. Передача опыта и знаний интеллектуальным роботам о техно-умвельтах осуществляется с помощью системы *технокодов*, которые передают информацию о выполняемых действиях и знаниях, получаемых из взаимодействия робота и техно-умвельта. Технокоды являются аналогом нейродинамических кодов человеческого мозга — функциональных состояний. Другой аналогией может быть универсальный прикладной программный интерфейс (API), который позволяет передавать разнородную информацию в универсальном формате.
4. Обработка технокодов, поступающих от различных техно-умвельт-машин выполняется с помощью специализированной интеграционной шины, связующей все четыре техно-умвельта в единое тьюринговое пространство.

Данная архитектура, основанная на независимо работающих умвельт-машинах и технокодах, которые передают информацию между ними, вполне способна эффективно управлять действиями интеллектуальных роботов, воспринимать мир и действовать во всех четырёх видах техно-умвельтов. Помимо этого, интеллектуальные роботы, обладающие данной архитектурой, имеют больше предпосылок для становления сознания (обобщённо — сложных психических феноменов), чем системы, действующие лишь в одном типе техно-умвельтов. На современном этапе развития науки использование результатов феноменологии сознания и достижений в области нейронаучных исследований является важным условием построения новых посттьюринговых когнитивных архитектур для создания интеллектуальных роботов и Общего искусственного интеллекта.

Следуя описанному выше подходу, целевая посттьюринговая архитектура интеллектуального робота состоит из четырёх умвельт-машин и набора технокодов, способных обеспечить эффективное, автономное функционирование робота во всех техно-умвельтах.

Остановимся подробнее на вопросе возможной эмерджентности сознания в интеллектуальном роботе, обладающем описанной целевой когнитивной архитектурой. Функционалистская парадигма в исследованиях мозга предполагает, что все ментальные состояния коррелируют с физическими состояниями нейронов. Соответствующие сборки нейронов должны быть эмпирически идентифицированы с помощью наблюдательных и измерительных приборов. Например, при «чтении мозга» можно идентифицировать скопления активных клеток, связанные со словами и соответствующими объектами. Один нейрон не имеет решающего значения и может отличаться у разных людей. Существуют типичные схемы распределения с нечёткими формами, которые представлены в компьютерном моделировании¹⁹⁷. Исследования мозга все ещё далеки от наблюдения за деятельностью каждого нейрона в мозге. Тем не менее сумма уже проведённых исследований, а также модель нейродинамических кодов психических явлений по крайней мере позволяет объяснять часть сложных психических состояний. В этой модели сознательные состояния означают, что люди рефлексируют свою деятельность. Самосознание реализуется с помощью дополнительных областей мозга, отслеживающих нейронные корреляты. Как отмечал Д. И. Дубровский, возникновение субъективной реальности как феномена может быть связано с тонким взаимодействием моторики и психики¹⁹⁸.

В связи с этим встает неизбежный вопрос: обладает ли предложенная посттьюринговая архитектура для интеллектуального робота возможностью для создания эмерджентности, превышающей возможности, которые может

¹⁹⁷ Дубровский Д. И. Проблема «сознание и мозг»: Теоретическое решение. М.: Канон+, 2015. 208 с.

¹⁹⁸ Дубровский Д. И. Субъективная реальность как предмет междисциплинарного исследования // Познание и сознание в междисциплинарной перспективе. Часть 1. М.: ИФ РАН, 2013. С. 16; Риццолати Дж., Синигалья К. Зеркала в мозге. О механизмах совместного действия и сопереживания / Пер. с англ. О. А. Куракова, М. В. Фаликман. М., 2012. С. 97.

дать среда любого из четырёх техно-умвельтов? Ведь в ходе биологической эволюции эффективные информационные и логистические процедуры были разработаны без какого бы то ни было символического представления в компьютерных моделях.

Нам представляется, что следует дать позитивный ответ на этот вопрос, так как каждый из техно-умвельтов даёт свои собственные свойства непредсказуемого поведения в уникальной среде. Оперирование сразу в четырёх техно-умвельтах требует разноуровневой рефлексии, что создаёт предпосылки для возникновения более сложных структур психики. Инженерная парадигма конструирования, основанная на работе С. Труна «Вероятностная робототехника»,¹⁹⁹ может иметь роль, подобную той, которую играют случайные мутации в биологическом процессе. Этот подход открывает возможность необходимого для эволюционного накопления опыта: с одной стороны, новые инженерные решения играют роль мутации, постоянного, полухаотического обновления конструкций роботов и их программного обеспечения, с другой — необходимость для машины действовать одновременно в четырёх техно-умвельтах играет роль фильтра, аналога естественного отбора, причем не по массе или другому количественному показателю, а по уровню сложности. При этом накопление информации, осуществляемое как роботом, так и конструкторским бюро, позволяет сохранять наиболее эффективные решения.

Предлагаемая диссертантом архитектура соответствует принципам постнеклассической рациональности, так как при подобном подходе, по выражению В. С. Степина, «вещь и процесс меняются местами: процесс становится первичен. Значительно расширяется понимание причинности: лапласовская причинность сохраняется, но она дополняется так называемой вероятностной причинностью»²⁰⁰. Далее нейронная самоорганизация может

¹⁹⁹ *Thrun S., Burgard W., Fox D.* Probabilistic robotics. Cambridge, MA: MIT Press, 2005. 480 p.

²⁰⁰ *Степин В. С.* Постнеклассическая рациональность и информационное общество //

вероятностным образом порождать сети, оперирующие в различных техно-умвельтах. Они будут как минимум эквивалентны сложным системам, моделируемым нелинейными дифференциальными уравнениями в создаваемых интеллектуальных роботах или системах искусственного интеллекта. Поскольку эти роботы решают задачи с определённой степенью сложности, то можно считать, что они обладают соответствующей степенью интеллекта. В ходе технической коэволюции человеческая деятельность интегрирована в социотехнические системы, решающие задачи высокой степени сложности.

В данном случае возникает своего рода человеко-машинный интеллект, который уже нельзя разделить ни на человеческий, ни на машинный. Это гибридный, техночеловеческий, робочеловеческий интеллект, который использует как знание человека о том или ином умвельте, так и знание робота о собственных техно-умвельтах, полученное или приобретенное им в реальности или в симуляциях. Человек не сводим только к научно-инженерному дискурсу, он участвует и в культурно-антропологическом, и в потребительском²⁰¹, поэтому машина для эффективного взаимодействия с ним также должна выходить за пределы научно-инженерного дискурса.

Сознание не тождественно интеллекту, и наличие сознания в любом случае — это не необходимое условие создания интеллекта. Интеллектуальная сеть, например, имеет определённую степень интеллекта и автономии, но не обладает сознанием. Но сеть, оперирующая в четырёх техно-умвельтах, постоянно решает сравнительно сложные задачи в том числе по рефлексии и саморефлексии. И тем формирует предпосылки для возникновения аналога человеческого сознания. Если рассматривать сознание как особое состояние мозга или искусственного интеллекта, которое в определённой степени

Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (17–18 марта 2016 г., г. Москва, философский факультет МГУ им. М.В. Ломоносова). М.: ИИнтелл, 2017. С. 59—70.

²⁰¹ Розин В. М. Понятие и современные концепции техники. М.: ИФ РАН, 2006. С. 37.

проявляется в процессе эволюции, как ответ на возникающую сложность умвелта, окружающего биологическую или техническую систему, — эволюционное накопление опыта в итоге должно приводить к осознающему себя искусственному интеллекту. Одновременное проектирование, программирование подобной машины сейчас невозможно. Но предлагаемая архитектура позволяет создавать сравнительно устойчивые конфигурации программного обеспечения, которые позволят пошагово наращивать сложность рефлексии и потом смогут стать единой теорией Общего искусственного интеллекта и робототехники, в которой человеческий интеллект является лишь составной частью.

3.2. Экспериментальный интеллектуальный робот «Э.ЛЕНА» как демонстрация применимости посттьюринговой методологии

Хотя уровень технологий 1940-х делал малоперспективными работы по совершенствованию «физических» роботов, А. Тьюринг не обошел молчанием этот вопрос. В отчете 1948-го он, на основе имевшихся в те годы механизмов, создал образ самодвижущейся интеллектуальной машины. Её воплощение он посчитал преждевременным, потому как «блуждающая по полям» Англии машина станет «несомненно создавать угрозу для окружающих»²⁰². Ещё С. Харнадом было показано, что именно под воздействием этих ограничений основания методологии А. Тьюринга редуцировались до символической, вербальной речи, а именно до переписки через телетайп. Другие формы общения человека и машины были бы либо невозможны, либо малоинформативны (хотя А. Тьюринг не отказался бы от использования искусственной плоти, которую упоминал в одной из работ²⁰³). Именно поэтому выбор А. Тьюринга пал на техно-умвелт *виртуальное-вербальное*, что породило бесчисленные попытки создания различных систем

²⁰² Turing A. M. Intelligent Machinery // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 508.

²⁰³ Turing A. Computing machinery and intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 553.

искусственного интеллекта, существование которых ограничено лишь единственным из четырёх возможных техно-умвельтов.

С глубокой древности люди находили творческие способы передачи изображений, будь то автопортреты или образы фантастических существ. Однако только изобретение фотографического процесса, электронно-лучевой трубки и радиосвязи сделало возможным мгновенную, качественную передачу изображения и звука от устройства к устройству.

Первоначально, с начала XX века было два направления записи, сохранения, передачи и воспроизведения аудиовизуальной информации: реалистичный (хотя и художественно преобразованный) и графический. Примером первого подхода является кинематограф, в котором фотография (реалистичный образ окружающего мира) используется для передачи двумерного (монохромного или цветного) или трёхмерного изображения. Примером второго подхода является мультипликация, в которой нарисованные (человеком или компьютером) изображения используются для создания виртуальной реальности. Оба подхода имеют общую идею — обман зрителя²⁰⁴. Однако технические и художественные средства используются для этого совершенно различные.

Оба направления передачи смысла используются для создания иллюзии реальности, которая в основе имеет текст, и «которая понята как текст»²⁰⁵.

Но следует учесть, что все предыдущие поколения динамических изображений, начиная от чёрно-белых мультфильмов Диснея и завершая сложной компьютерной графикой, были лишь мимезисом реальности, причём семиотический план окружающего мира и общения людей являлся основой творчества. Художники, моделирующие в «3D Maya» одуванчики и тракторы, создают новую виртуальную реальность, для которой эталоном качества остаются работы классических живописцев — взгляд художника должен

²⁰⁴ Аронсон О. В. Кино и философия: от текста к образу. М.: ИФ РАН, 2018. С. 8.

²⁰⁵ Там же. С. 10.

влиять на впечатление зрителей, которые узнают в изображении что-то привычное.

Совершенно другой подход, который основан на сочетании графического и реалистического, появляется сейчас благодаря новым технологиям. Компьютерная графика создаёт не виртуальную альтернативу окружающему миру, а технологически продолжает и дополняет его: исходное изображение (портрет или пейзаж) дополняется текстовыми характеристиками, после чего машина создает новое динамическое изображение. Эта картина выглядит абсолютно реалистично, но при этом может никогда не существовать в действительности. Уместной иллюстрацией здесь выступит известная история античной эпохи: соревнование художников Зевксиса и Парразия: первый нарисовал виноградную гроздь, которую слетелись клевать птицы, но второй нарисовал покрывало на картине, которое обмануло Зевксиса, думавшего, что покрывало надо отдернуть, чтобы рассмотреть само изображение²⁰⁶, то есть иллюзия была совершеннее и обманывала уже человека.

Проводя аналогию, можно уподобить мастерство Парразия обработке реального изображения компьютерными нейронными сетями, что позволяет дополнить или даже восполнить реальность. Тогда мастерство Зевксиса — это всего лишь копирование оригинала с предельно возможной точностью.

То есть для художника компьютер становится новой кистью, но не «карандашом», как писал К. Поппер²⁰⁷.

Возможности дополненной реальности или компьютерной графики сами по себе не рассматриваются в данной работе. Необходимо сопоставить способы коммуникации человека и компьютера: будут ли электроника и программы надёжным, объективным инструментом или же способом (medium) чтобы общаться с другими людьми посредством сообщений

²⁰⁶ Давыдова Л. И., Конькова Г. И., Чубова А. П. Античные мастера. Скульпторы и живописцы. Л.: Искусство, 1986. С. 51—52.

²⁰⁷ Поппер К. Логика и рост научного знания. М.: Прогресс, 1983. С. 518.

(message)? Как установить ту грань, за которой люди потеряют возможность различать сообщения компьютера и человека? Сходство сообщений от двух корреспондентов будет достаточным для обмана. Известный афоризм М. Маклюэна о том, что «средство коммуникации и есть сообщение»²⁰⁸ можно расширить, сформулировав утверждение, что «машина и есть средство коммуникации». В 1960-х, когда М. Маклюэн опубликовал свою книгу, образ целиком виртуального диктора, воплощённой компьютерной программы, был технологически нереалистичным. Сегодня же это реализованный проект.

Лаборатория робототехники Сбербанка смогла начать применение первого цифрового русскоязычного диктора «Э.ЛЕНА» (Электронная ЛЕНА) еще в 2019 г. Достигнута генерация реалистичного образа телеведущей в режиме реального времени²⁰⁹. Нельзя сказать, что подобная технология принципиально уникальна: искусственные нейронные сети широко применяются для улучшения графических изображений. Распространены программы, которые позволяют убрать эффект «красных глаз», обогатить колористику снимка, создать эффект «боке» с помощью размытия фона. Эти изменения эстетически приятны и настолько легко воспринимаются потребителями, что ясное объективное восприятие реальности даже получило обозначения (тег) *#nofilter*. Но принципиальная новизна технологических возможностей состоит в дополнении реальности или же виртуальном реконструировании объектов, причём в режиме реального времени. Объекты на экране подчёркнуто лишены привычных дефектов компьютерной графики или эстетически гипертрофированных черт, они кажутся реальными, документальными, порой просто знакомы потребителю с детства. Одновременно показано использование этих объектов, которое в реальности никогда не происходило и случиться принципиально не может. *Deepfake* —

²⁰⁸ Маклюэн М. Понимание Медиа: внешние расширения человека. М.: Гиперборей, 2007. С. 11.

²⁰⁹ Елена. Цифровой двойник телеведущей Сбербанка. URL: https://www.tadviser.ru/index.php/Продукт:Елена_Цифровой_двойник_телеведущей_Сбербанка (дата обращения: 21.07.2020).

обозначение новой социально-технологической практики, которая пока, к сожалению, имеет преимущественно негативное либо развлекательное использование. Между тем, принципиальная проблема отчуждения образа и актрисы путём оцифровки уже поднималась в фантастическом фильме («Конгресс», 2012)²¹⁰.

Попытку разработать теледиктора, который был бы двойником реального человека, предприняла компания *Sogou* (заказ агентства Xinhua), готовое платформенное решение они представили в 2018-м²¹¹. Сбербанк через несколько месяцев представил иное решение: русскоязычный теледиктор, который может озвучить любой текст в целиком автоматическом режиме. Подобные модели могут широко использоваться в корпоративном телевидении. Фактически уже сложилась практика использования Э.ЛЕНЫ: произведены десятки новостных сюжетов, а сотрудники и клиенты Сбербанка сотни тысяч раз видели ролики и узнавали новости из уст цифрового диктора. Инженерный коллектив Лаборатории робототехники Сбербанка под руководством диссертанта получил патент на изобретение на эту технологию²¹².

Робот «Э.ЛЕНА» является виртуальным гуманоидом, наделённым мимикой, может использовать естественный язык, полностью поддерживая русский язык (виземы, фонемы), является автономным, т. е. не требующим действий оператора при выполнении базовых функций, обладающий собственной персоной в зависимости от персоны чат-бота и имеющий первичную возможность для обучения изменениям внешнего мира (в работе

²¹⁰ Конгресс (The Congress, реж. А. Фольман, Й. Гудман, 2013). URL: [https://ru.wikipedia.org/wiki/Конгресс_\(фильм,_2013\)](https://ru.wikipedia.org/wiki/Конгресс_(фильм,_2013)) (дата обращения: 22.06.2020).

²¹¹ Chinese state media's latest innovation is an AI female news anchor // CNN Business. URL: <https://qz.com/1554471/chinas-xinhua-launches-worlds-first-ai-female-news-anchor/> (дата обращения: 12.08.2020).

²¹² Ефимов А. Р., Гонноченко А. С., Владимиров М. А. Способ и система для создания мимики на основе текста // Патент на изобретение RU 2 723 454. 2020.

реализовано распознавание лиц собеседников робота). Впервые в литературе представлена архитектура подобного робота²¹³.

Для синтеза голоса Э.ЛЕНЫ используется программное обеспечение для преобразования произвольного текста в речь с помощью искусственных нейронных сетей, обученных на специализированном множестве фонетических данных. В результате получается непрерывный нейросетевой синтез речи. Далее действует несколько искусственных нейронных сетей, обученных с использованием видеоматериалов и 3D-сканов актрисы (для более эффектного сочетания голоса и лица), которые создают мимику диктора.

В Э.ЛЕНА преобразование текста в видео-речи диктора с отдельным созданием мимики проходит целиком автоматически. После чего для устранения ошибок проводится контрольная обработка, которая также целиком автоматизирована. Видео — достаточно реалистичное, чтобы обмануть многих пользователей, — будет целиком готово после того, как по согласованию с заказчиком выберут фон. Комплекс не разделён на отдельные модули, для обращения с каждым из которых нужен специалист, а целиком интегрирован в единую платформу, снабжённую компонентами искусственного интеллекта.

Сегодня служба корпоративного телевидения ПАО «СБЕРБАНК» является единственным пользователем сервиса, который позволяет создавать готовый видеофрагмент на основе текста. Данная программа позволила сгенерировать более сотни разнообразных новостных сюжетов. Чтобы определить способности Э.ЛЕНЫ вводить в заблуждение пользователей своей внешностью, был проведён специальный опрос. Он проводился в сети «Одноклассники», количество опрошенных составило 22 тысячи человек, причём мнения разделились. В ответ на вопрос о том, в каком из двух

²¹³ *Ефимов А. Р.* Технологические предпосылки неразличимости человека и его компьютерной имитации // Искусственные общества. 2019. Т. 14. С. 74–80. DOI: 10.18254/S207751800007645-8. URL: <https://artsoc.jes.su/s207751800007645-8-1/> (дата обращения: 19.07.2020).

видеороликов показан живой, а в каком виртуальный диктор, — более четверти опрошенных дали неправильный или неуверенный ответ.

Естественно предположить, что применение данных технологий будет расширено, а качество улучшено. Ими смогут пользоваться коммерческие фирмы или даже обычные пользователи.

Социокультурные особенности, контекст взаимодействия людей, обстоятельства места и времени, не говоря уже о целях участников (и взаимных догадках о целях собеседников) — это важнейшая составляющая диалога. Взаимное понимание смысла коммуникации лишь на основе вербально выраженной интенциональности, порой, просто недостижимо.

Язык, контекст, культура — вот три необходимых условия выражения смысла, при обеспечении которых можно в общем виде решить задачу взаимного понимания машины и человека. Поэтому многолетняя практика редукции диалога человек-машина лишь к вербальной составляющей, представляется принципиально недостаточной, ограниченной, так как утрачивается возможность передачи смысла с помощью невербальной информации. Прямолинейное, буквальное значение слов не может передать все три модальности смысла — культуру, язык и контекст. Они присутствуют в диалоге не только как буквальное понимание значения слов его участниками, но и в виде имплицитного культурного знания, которое Д. Эверетт метафорически именуется «тёмной материей». Смысл слов и выражений сам по себе дополняется жестами, окружающими предметами или даже общими ощущениями участников диалога, причем «тёмная материя» может полностью менять смысл произнесённых с той же интонацией предложений.

В случае ЭЛЕНЫ более высокий уровень иллюзии достигается благодаря одновременному действию программы в двух техно-умвельтах: *виртуальном-вербальном* (слова диктора) и *виртуальном-невербальном* (мимика и жесты диктора).

В художественных артистических практиках этот эффект давно известен: актёры по методике К. С. Станиславского должны полусотней

способов, с различной интонацией произнести фразу «завтра вечером» (М. Долар²¹⁴). В любом очном диалоге используется мимика лица, интонация и жесты. Они демонстрируют личный жизненный опыт собеседника и особенности его биографии, выраженные в социальном статусе и других особенностях его культуры. По выражению Д. Эверетта «язык никогда не выражает всего, культура заполняет эти пропуски»²¹⁵. Увы, сложившаяся традиция тьюрингового подхода требует общаться с машиной с помощью фактически телетайпа, пусть и более современного, в форме лаконичных текстовых сообщений. «Тёмная материя» коммуникации отсекается, ведь жесты и мимика игнорируются, интерпретация сообщения по Д. Эверетту становится затруднённой. Создание Э.ЛЕНЫ открывает путь к техногенному исследованию этой «темной материи»: получен испытательный образец, в котором можно совмещать однозначность, заданность жестов и мимики теледиктора с широким охватом аудитории. Пропуски и неясности в «телетайпных» диалогах можно исследовать с помощью невербального общения.

Разумеется, Э.ЛЕНА не станет бэконовским «*experimentum crucis*», окончательно опровергающим концепцию функционализма: как справедливо заметила Е. А. Мамчур, само существование таких экспериментов — дискуссионный вопрос²¹⁶. Проблема невербального диалога компьютера и человека не затрагивается в львиной доле научных и популярных работ о проблемах искусственного интеллекта. Исключением можно назвать советский научно-популярный фильм «Кто за стеной?» 1977 года, в котором собеседники (среди которых находился компьютер) судей-людей были

²¹⁴ Долар М. Голос и ничего больше / Пер. с англ. А. Красовец. М.: Изд-во Ивана Лимбаха, 2018. С. 85.

²¹⁵ Эверетт Д. Как начинался язык. История величайшего изобретения. М.: Альпина Нон Фикшн, 2019. С. 384.

²¹⁶ Мамчур Е. А. Существует ли в современном естественно-научном познании бэконовский «*experimentum crucis*»? // Vox. 2019. № 26. С. 141–153. URL: <https://vox-journal.org/html/issues/480/502> (дата обращения: 23.07.2020).

представлены в виде изображений своих лиц на телеэкранах²¹⁷. Сложившаяся традиция сформирована не просто стилистикой работ А. Тьюринга или верификационными требованиями неопозитивизма, но вполне конкретным ограничением, «стеной Тьюринга»: игра в имитацию должна проводиться с помощью записок, которые нивелируют сопутствующие факторы, форму сообщения, и значимым там должно быть лишь их содержание. Однако форма и содержание не могут быть полностью окончательно разделены: технологически неизбежное в 40-е годы XX века деление не только стало ограничением в исследованиях искусственного интеллекта, но может быть преодолено на современном уровне техники.

Функционалистская методика А. Тьюринга, основой которой можно назвать «тест Тьюринга» может быть расширена благодаря инструментам дополненной реальности, цифровым аватарам и теледикторам типа Э.ЛЕНА. Обсуждение возможностей, которые дают роботы типа Э.ЛЕНА, поднимает другой вопрос: где находится предел имитации способностей человека по использованию мимики и жестов, а также всего спектра запахов и всех шумов, которые производят человеческие организмы? Представляется, что насколько невербальные коммуникации во всём их многообразии станут доступны машинам, настолько увеличатся, по сравнению с человеческими, их арсеналы локутивного акта. Если подобные технологии окажутся достаточными для имитации всего спектра сообщений, который может получить человек, то, фактически, для определения статуса собеседника у нас останется лишь эмпирический подход: с помощью физического воздействия определить, участвует ли в диалоге живой человек или набор микросхем и программных кодов.

Можно ли считать Э.ЛЕНУ новым инструментом, использование которого позволит решать проблемы искусственного интеллекта или можно возразить, что данная разработка лишь новый вид компьютерной графики и не

²¹⁷ Кто за стеной? (Реж. С. Райтбурт, 1977). URL: <https://www.youtube.com/watch?v=ECvsD4b0JIU> (дата обращения: 04.05.2020).

имеет отношения к исследованиям искусственного интеллекта. К этому возражению можно привести контрдовод: данная технология качественно меняет проведение классического теста Тьюринга. Дело не только в том, что игроки в имитацию могли бы эффективнее и эффективнее обманывать судью с помощью цифрового аватара. Происходит объединение формы и содержания сообщения, что требует от человека или компьютера использования на порядок более широкого культурного багажа, причём в режиме реального времени. Распространение видеосвязи, как инструмента общения среди молодежи, легко объясняется по Д. Эверетту: для понимания смысла невербальные коммуникации могут иметь значение не меньшее, чем сама речь. При жизни А. Тьюринга проекты по реализации аналога Э.ЛЕНЫ были бы чистой утопией. Сейчас это значимая, но не экстраординарная технологическая новация, которая быстро получает распространение и в любом случае даст богатый эмпирический материал для описания взаимодействия человека и машины. Поэтому теоретическое и философское осмысление этого тренда представляется чрезвычайно актуальным.

На основе вышеперечисленного можно сформулировать дальнейшие требования к совершенствованию Э.ЛЕНЫ:

- возможность предварительного задания культурных контекстов. Например, торжественного голоса и воодушевленной мимики при озвучивании особо важных сообщений. Естественно, задавать все подробности непродуктивно, требуется какая-то доступная оператору шкала «торжественности», по которой можно просто передвинуть ползунок;

- возможность Э.ЛЕНЫ цензурировать или даже стилистически обрабатывать заданный текст. Например, удалять пропущенные оператором непристойности и устранять тавтологии. Возможности перевода ограниченного набора протокольных иностранных фраз, особенно с малоизвестных в России языков;

- возможность хотя бы частичного смыслового анализа поступающих текстов и видеоматериалов, причём реакция на это должна быть как минимум

в двух техно-умвельтах. Если видеосюжет выводится на задний план или в кадре появляется другое действующее лицо, то требуется сместиться так, чтобы не заслонять собой содержательные фрагменты изображения, дать собеседнику место в кадре. Также *виртуальное-невербальное* взаимодействие должно включать распознавание Э.ЛЕНОЙ истерической жестикуляции собеседников. Если речь идёт о несчастных случаях, о трагедиях, репортаж о которых передается в прямом эфире и оператор заведомо не успеет изменить культурологические настройки Э.ЛЕНЫ, желательно, чтобы она сама убирала из речи радостные коннотации — это *виртуальное-вербальное*;

– в идеале отслеживание Э.ЛЕНОЙ текущего контента по жестикуляции, мимике и речевой культуре референтного набора дикторов.

Да, нынешний уровень развития техники далёк от фантастической идеи А. и Б. Стругацких о том, чтобы «модели на базе собственных дублей умел создавать любой сотрудник, защитивший магистерскую диссертацию»²¹⁸. На сегодняшний день создание видеодвойника — весьма трудоёмкий процесс, который невозможен без привлечения сотрудников самых разных компетенций. Результат несовершенен, уступает по своим возможностям актрисе — например, пока цифровому диктору совершенно недоступен танец. Но в качественном смысле рубеж создания «видеодвойника» преодолён и дальнейшее усовершенствование технологии пойдёт более быстрыми темпами. Человеческая профессия теледиктора, а частично и актёра, может уйти в прошлое, сменившись специализацией компьютерных программ. И уже эта перспектива демонстрирует, что улучшенное эмоциональное общение человека и цифрового двойника никак не гарантирует единства их интересов. Антропоморфность машин — во многом внешняя, потому что управление самыми совершенными андроидами может осуществляться удаленно, что не гарантирует гуманизацию технологии. Прогресс в создании Общего искусственного интеллекта требует дополнительных предосторожностей.

²¹⁸ Стругацкий А. Н., Стругацкий Б. Н. Понедельник начинается в субботу: Сказка для научных работников младшего возраста. М.: Детская литература, 1965. С. 126.

3.3. Мироззренческие перспективы постгьюринговой коммуникации людей и роботов

Ю. Харари в рецензии на книгу М. Тегмарка про искусственный интеллект отметил, что «если вам рассказывают сценарий о мире в 2050 г., который вам кажется научной фантастикой, то, вероятно, это неверно; но если вам рассказывают сценарий о мире в 2050 г., и он не похож на научную фантастику, это определено неверно»²¹⁹. Высказываются гипотетические соображения относительно будущего взаимодействия людей и роботов при условии того, что роботы станут технологически неразличимы от людей в ряде сценариев, которые могут иметь значительное влияние на нашу жизнь и жизнь наших потомков. Подобные сюжеты давно и широко обсуждаются в футурологической и философской литературе (В. Г. Буданов)²²⁰. Здесь обычно сочетаются разноплановые вопросы. Одно дело, когда речь идёт о неразличимости человеком-наблюдателем поведения человека и робота в некоторых конкретных случаях, другое — когда во всех случаях, т.е. речь тогда идёт фактически о тождественности человека и робота. Но это означает, что в ходе антропотехнологической эволюции возник новый вид человека, постчеловек, о котором пишут трансгуманисты. Однако их позиция является слишком дискуссионной. Мы имеем в виду неразличимость в первом случае (то, что можно как раз назвать «технологической неразличимостью»). Но и в такой плоскости часто ставится глобальный вопрос о новом этапе развития робототехники, когда робот обретает все большее число человеческих способностей, может достигнуть уровня человека и затем превзойти его (а человек тем временем развивается в своем биологическом облике).

Для реалистичной оценки форм взаимодействия человека и робота, для устранения распространённых в публицистической литературе сюжетов о неизбежном уничтожении людей конкурирующими машинами, было

²¹⁹ Харари Ю. Н. 21 урок для 21 века. М.: Синдбад, 2019. 416 с.

²²⁰ Буданов В. Г. Новый цифровой жизненный техноуклад — перспективы и риски трансформаций антропосферы // Философские науки. 2016. № 6. С. 47–55.

уточнено понятие **Общего искусственного интеллекта** — это способность робота (или иной технической системы) обучаться и действовать совместно с человеком или автономно в любой области, но лучше, чем специалист в этой области, достигая поставленных целей во всех четырёх техно-умвельтах при ограничении потребляемых роботом ресурсов.

Моделирование и программирование таких свойств роботов, которые бы отвечали нашим юридическим и этическим принципам, полностью исключали бы их агрессивность и «недружественные интенции», вероятно, потребует создания «виртуальных людей» — программ, которые будут эмоционально отождествлять себя с людьми, обладать чертами человеческого самосознания и самости. Рассматриваются вопросы использования таких роботов в областях образования, производства и развлечения. Это вопросы о создании автономных виртуальных учителей, основанных на учёте персональных, в том числе генетических характеристик учащихся или, например, вопросы авторского права на произведения искусства, создаваемые роботами, включая продукцию на основе цифровых двойников известных личностей, как ныне живущих, так и ушедших. Ставится и обсуждается вопрос о том, как провести границу между человеческим и технологическим интеллектом, если будет достигнут уровень Общего искусственного интеллекта.

Всё это свидетельствует о возрастающей роли философских и методологических подходов в решении задач развития интеллектуальной робототехники и искусственного интеллекта в целом.

Рассмотрение возможных вопросов и гипотетических сценариев мы начнем с анализа, сконструированного А. Ю. Алексеевым мысленного эксперимента «девушка по переписке». Однако не в форме простого повторения, а в форме синтеза с постъюринговой методологией, предложенной в данной работе.

«Девушка по переписке». Молодой человек ищет спутницу жизни и обращается к службе интернет-знакомств. Задает традиционные параметры запроса — возраст, внешность, образ жизни. Получает от сайта имена и

фотографии ряда кандидаток. Его выбор падает на Элизу. Электронное общение длится несколько месяцев и шансов для ошибки, казалось бы, не остаётся. Наконец, молодой человек назначает свидание и получает согласие. Приезжает с традиционным подарком, например, букетом цветов. Однако же никто к нему на встречу не приходит. Естественно, отправляет письмо с возмущениями. В ответ: «Прости, милый! Я — Электронная Лиза, компьютерная программа»²²¹.

Если вынести за скобки естественную глупость молодого героя и условность этого мысленного эксперимента, то, по мнению А. Ю. Алексеева, этот сценарий мог бы быть адекватным свидетельством о прохождении искусственным интеллектом теста Тьюринга в соответствии с ранее раскрытой тьюринговой методологией. Молодой человек, выполнивший роль судьи-наблюдателя, слишком доверился «наивной феноменологии» восприятия другого, перенеся на «электронную Лизу» антропоморфную форму, дорисовав в своем воображении «её небесные черты».

Предполагая, что развитие искусственного интеллекта и компьютерных технологий пойдёт по линейному пути сохранения существующих тенденций к улучшению методов обработки речи с помощью компьютерной лингвистики, которые предсказывал еще А. Тьюринг, А. Ю. Алексеев считал, что «в недалёком будущем воплотится и сценарий “Девушка по переписке”»²²².

Исходя из сказанного в Главах 1, 2 и 3 мы предполагаем, что даже если этот сценарий и может воплотиться, он уже не имеет никакого смысла, так как «переписка через стену Тьюринга» лишает машины (какую бы функцию она не выполняла) возможностей взаимодействия с человеком. Прежде чем продолжить, предложим свой вариант мысленного эксперимента

²²¹ Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. М.: ИинтелЛЛ, 2013. С. 15.

²²² Там же. С. 16.

А. Ю. Алексеева с учётом современных реалий и предполагаемых тенденций развития искусственного интеллекта²²³.

«Девушка по переписке — 2025». Скучающий молодой человек ищет в приложении для взрослых на смартфоне спутницу жизни. Приложение уже имеет все необходимые данные о молодом человеке, включая время, на которое он задерживается на фото с красивыми девушками, все его поисковые запросы, посещения концертов, музеев и фотографии из путешествий в отпусках, данные фитнес-трекеров, средний размер чека в ресторанах и т. п. Искусственный интеллект предлагает молодому человеку выбор параметров запроса, даёт свои предложения и рекомендации на основе предоставленной им самим ранее информации, а также предлагает включить функцию «улучшения профиля», которая позволяет автоматически формировать ответы партнёру всего лишь за 500 руб./месяц. С точки зрения искусственного интеллекта этот шаг является пустой формальностью, так как 99% молодых людей с аналогичными запросами соглашались сразу же с рекомендованными параметрами, но молодому человеку хочется сохранять иллюзию свободного выбора, поэтому он отказывается от рекомендаций и улучшений. Сайт выдаёт перечень имен. Начинается переписка с девушкой по имени Лена. Переписка длится пару дней²²⁴! С ней можно беседовать обо всём: политика и экономика, зарплата и отдых, кухня и литература, вдохновение и секс. Играют в ролевые онлайн игры. Молодой человек назначает встречу в кафе, которое, судя по всему, нравится им обоим.

В назначенные день и время они оказываются в этом кафе. Зная предпочтения Лены, молодой человек сделал общий заказ. В ожидании заказа, он начинает рассказывать историю из новой серии их любимого комедийного сериала, но Лена не понимает шутку. Она в ответ ему говорит о своем новом

²²³ Впервые представлен на IV Международной конференции в области L&D (обучения и развития) «Больше чем обучение: как добиться изменения поведения?», 24–26 октября 2019 г. URL: <https://conference.sberbank-university.ru/> (дата обращения: 03.02.2020).

²²⁴ Очевидно, значительное уменьшение социальной дистанции и времени на принятие решений.

автомобиле, но молодой человек очень расстраивается, так как является категорическим сторонником зеленого образа жизни и считает очень вредным желание людей иметь автомобили. Молодой человек пытается рассказать о своем котёнке, но тут же понимает, что Лене нет никакого дела до кошек. Словом, что-то не так, хотя девушка вроде бы такая, как на фотографии — светлая, с умными глазами. Официант приносит заказ. Воспользовавшись паузой, молодой человек заглядывает в профиль Лены и видит, что в отличие от него, Лена — платный пользователь, который включил «улучшения, тип Э.ЛЕНА», использующие автоответы. Всё стало сразу понятно.

Свидание закончилось. Сидя в такси, молодой человек снова пишет в приложение о том, что ему немного не по себе, но Лена ему тут же отвечает фразой из его любимого сериала, а также шлет фото котика, которого она подобрала на улице. Переписка продолжается.

Суть посттьюринговой методологии в том, что **стены Тьюринга** между машиной и человеком более не существует ни при оценке полезности и рациональности действий машины, ни при реальном взаимодействии. Искусственный интеллект «Э.ЛЕНА» в приведённой выше модификации мысленного эксперимента является для человека более близким и понятным эмоционально, он «более человечен, чем сам человек»²²⁵. Именно это является ключевой характеристикой посттьюринговой методологии, в которой более нет места слепому сравнению функциональных возможностей и практической рациональности человека и машины. Машина (интеллектуальный робот) действует совместно с человеком, а человек знает, что он взаимодействует с машиной, потому что вместе с машиной ему лучше, надёжнее, практичнее и интереснее.

С одной стороны, машина-напарник будет подталкивать человека к деградации, открывая «возможность замены собственных мыслительных

²²⁵ Дик Ф. Мечтают ли андроиды об электроовцах / Пер с англ. М. Пчелинцев. М.: Fanzon, 2016. 640 с.

усилий готовой информацией»²²⁶. Но, с другой стороны, молодому человеку, если он не желает стать буквально наркоманом, зависимым от чисто электронного образа, требуется улучшать себя, соответствовать идеальному образу реальной девушки: «говоря о биотехнологическом инхенсменте, мы должны иметь в виду также и инхенсмент самого субъекта»²²⁷.

Данное ранее определение Общего искусственного интеллекта в контексте действия робота или иной машины, наделённой таким интеллектом, приобретает особый смысл, так как среду можно определить как техно-умвельт. Попробуем заново определить Общий искусственный интеллект с учетом вышесказанного.

Общий искусственный интеллект — это способность робота (или иной технической системы) обучаться и действовать *совместно с человеком или автономно в любой области*, но лучше, чем специалист в этой области, достигая поставленных целей во *всех четырёх техно-умвельтах* при ограничении потребляемых роботом ресурсов.

Четыре техно-умвельта, введённые и описанные ранее, взятые в своем целостном единстве, являются аналогом научной трансдисциплинарности, так как предполагают охват всего разнообразия видов взаимодействия машины, выступающей как субъект, и человека — от физического взаимодействия с помощью вербальной коммуникации до виртуального общения с помощью невербальных, неречевых видов коммуникаций.

Исходя из этого определения, необходимо рассмотреть возможности, которые открываются перед интеллектуальной робототехникой в случае преодоления технологических барьеров, упомянутых в параграфе 1.2.

²²⁶ Лекторский В. А. Философия. Познание. Культура. М.: Канон+, 2012. С. 174.

²²⁷ Инхенсмент — технологии, способствующие достижению счастья. — Тищенко П. Д. Биотехнологии инхенсмента: на пути к третьей утопии? // Рабочие тетради по биоэтике. Вып. 20: Гуманитарный анализ биотехнологических проектов «улучшения» человека: сб. науч. ст. / Под ред. Б. Г. Юдина. М.: Изд-во Московского гуманитарного университета, 2015. С. 28.

Технологии, развивающие виртуальных людей, подобные описанной ранее в параграфе 3.2 Э.ЛЕНЕ, могут достичь значительного совершенства и распространения в случае преодоления барьеров, связанных с обработкой естественных языков, значительного увеличения доступных для роботов вычислительных мощностей. Программы, которые будут создавать и совершенствовать таких виртуальных людей могут обладать теми же характеристиками, что и люди, живущие в реальном физическом мире. В работе Д. Бурдена и М. Савин-Бадена²²⁸ выделяется три градации таких виртуальных личностей: виртуальные гуманоиды, виртуальные люди и виртуальные сапиенсы. «*Виртуальные гуманоиды*» — это виртуальные люди, которые в лишь ограниченной степени представляются людьми и которые могут выражать некоторые виды поведения, эмоций, шаблонов мышления, частично повторяя автономию и взаимодействие физического человека. «*Виртуальные люди*» — это программы, которые представляются как люди и могут имитировать поведение, эмоции, мышление, автономию и взаимодействие, смоделированные на основе реальных физических возможностей человека. «*Виртуальные сапиенсы*» — это сложные *виртуальные люди*, которые достигают схожих уровней представления, поведения, эмоций, мышления, автономии, взаимодействия, самосознания и внутренней самости, сходные с реальным человеком. Виртуальные гуманоиды, люди и сапиенсы могут обладать следующими характеристиками:

– осуществляют действия, которые проявляются в визуальной, аудио- и текстовой форме. К примеру, современные чат-боты могут общаться с человеком лишь с помощью текстовых сообщений, но виртуальные люди могут вести себя полностью подобно Э.ЛЕНЕ, демонстрируя антропоморфный облик;

– могут быть как имеющими телесный облик, так и не имеющими его. К примеру, виртуальный человек может общаться лишь в аудио- и текстовом

²²⁸ Burden D., Savin-Baden M. Virtual Humans. Today and Tomorrow. Boca Raton: CRC Press, 2019. С. 6.

формате, не имея телесной оболочки даже на экране компьютера. Именно таким является виртуальная героиня фильма «Она» («Her», реж. С. Джонз)²²⁹;

– виртуальные люди могут иметь как форму гуманоида, так и не гуманоида. Скажем, для детей это может быть любая форма сказочного героя (Колобок, Кот-в-сапогах и т. п.);

– виртуальные люди могут использовать широкий диапазон языков. В настоящий момент голосовые (виртуальные) ассистенты используют больше командный язык, чем естественный. Однако виртуальные сапиенсы смогут понимать язык человека не хуже, чем это делает обычный человек.

Виртуальные люди могут иметь три степени автономности. Первая, самая низкая степень автономности — полное управление оператором. Это так называемый аватар, т. е. по сути дела — виртуальный образ реального человека. Основой действий такого аватара является ряд т. н. игровых механик. Вторая, более автономная степень, — это частичная автономность. Виртуальный человек может совершать ряд действий в виртуальном мире, но нуждается в постановке задач или исправлении неточностей со стороны человека. Третья, высшая степень автономности, — полная автономность. Созданный виртуальный человек полностью независим от оператора и выполняет поставленные задачи самостоятельно. Такой степенью автономности будут обладать виртуальные сапиенсы.

Виртуальный человек будет демонстрировать все базовые эмоции, которые есть у любого человека: радость (довольство), удивление, печаль (грусть), гнев (злость), презрение, страх — уже сейчас эти эмоции хорошо описаны, и их воспроизведение алгоритмизировано²³⁰.

Виртуальный человек на уровне «виртуальный сапиенс» может обладать искусственно созданной личностью, имеющей уникальный характер.

²²⁹ Она (Her, реж. С. Джонз, 2013). URL: [https://ru.wikipedia.org/wiki/Она_\(фильм,_2013\)](https://ru.wikipedia.org/wiki/Она_(фильм,_2013)) (дата обращения: 02.08.2020).

²³⁰ *Eckman P.* Increase your emotional awareness and detect deception. URL: <https://www.paulekman.com> (дата обращения 18.08.2020).

В. К. Финн²³¹ выделял ряд характеристик, которыми мог бы владеть такой виртуальный сапиенс. В частности, возможно полностью реализовать способности, перечисленные в Главе 1, параграф 1.2 (1, 3, 4, 5, 6, 8, 9 и 10), включая способности к рассуждению, обучению и рефлексии.

Виртуальные сапиенсы в интерактивном режиме (т. е. совместно с человеком) смогут демонстрировать реализацию функций 2, 7, 12 и 13. В частности, обладать воображением и самосознанием.

Описанные выше характеристики виртуальных людей, сапиенсов открывают широкие возможности для различного рода применений, которые могут кардинально изменить нашу цивилизацию. В социальном программировании индивидов и социумов²³² будет достигнута качественно новая многосубъектность. При этом открывается возможность не следовать жёстко идеям трансгуманизма, а воспользоваться «спектром методологических регулятивов»²³³, гармонизирующих отношения людей и машин. Рассмотрим некоторые возможные следствия в ряде областей.

Образование. Возможно создание полностью индивидуализированного виртуального учителя, который взаимодействует со своим учеником по всему спектру обучения в полностью автономном режиме, предлагая ученику индивидуальную траекторию развития, основываясь не только на результатах ученика, но и на анализе ДНК, социальном окружении и личных особенностях характера. Естественно, этот виртуальный учитель с возрастом постепенно превращается в виртуального помощника, секретаря и консультанта. Поскольку такая программа будет обладать исчерпывающим набором информации о личности хозяина, то это поставит новые проблемы в защите личных данных, свободы и ответственности индивида.

²³¹ Финн В. К. Искусственный интеллект: Методология, применения, философия. М.: Красанд, 2018. С. 37–39.

²³² Шалак В. И. Социальное программирование // SocioTime / Социальное время. 2016. № 1 (5). С. 57.

²³³ Лисеев И. К., Петрова Е. В., Фесенкова Л. В., Хен Ю. В. Науки о жизни сегодня: философские инновации. М.: ИФ РАН, 2016. С. 54.

Возможно, уже при проведении научных исследований подобные программы смогут снизить барьер узкой специализации учёного, который описывает А. Л. Никифоров,²³⁴ — за счёт индивидуальной обработки справочной информации. И они же во многом решат проблему информационного перенасыщения, связанную с агрессивным навязыванием пользователю избыточной информации и односторонней подачи запрошенной²³⁵.

Производство. Возможно создание производственного предприятия, полностью лишённого управленческого или инженерного состава. На сегодняшний день мы видим примеры таких предприятий на основе таксомоторных парков и транспортных служб (UBER, Яндекс-такси и проч.), в которых у водителей нет никаких руководителей, кроме алгоритмов, которые ставят им задачи, выдают маршруты и выплачивают вознаграждение. Однако совершенно реально создать производство, в котором человеку остаётся лишь роль посредника между станком (им может быть программное обеспечение) и изделием (им может быть также любая услуга). Это поднимет вопрос о контроле этих предприятий, которые со стороны будут восприниматься в образах автономных техноценозов своеобразной техножизни.

Развлечения. Возможно создание полностью виртуального мира, населённого актёрами, которые уже ушли от нас. Но в этом мире актёры могли бы играть любую пьесу по желанию зрителя. Иными словами, зритель мог бы выбирать из уже готового набора пьес и постановок или создавать собственную постановку, приглашая на роли известных актёров прошлого. Виртуальные сапиенсы могут иметь «личины» актёров, искусственно восстановленные голоса и играть в пьесе. К примеру, в пьесе А. П. Чехова

²³⁴ Никифоров А. Л. Фундаментальная наука в XXI веке // Вопросы философии. 2008. № 5. С. 58–61.

²³⁵ Труфанова Е. О. Информационное перенасыщение: ключевые проблемы // Философские проблемы информационных технологий и киберпространства. 2019. Т. 16. № 1. С. 14.

«Вишневый сад» Раневскую могла бы сыграть Элина Быстрицкая, а Гаева — Андрей Миронов. Иннокентий Смоктуновский мог бы сыграть в такой пьесе Фирса. Технология deepfake уже демонстрирует подобное²³⁶. Но представим, что виртуальные сапиенсы будут развивать сюжет, обустривать некий виртуальный мир и продолжать «жизнь» в нём, пока заказчики в состоянии оплачивать расходы на электричество. Виртуальные вселенные многопользовательских игр станут индивидуальными.

Описанные выше возможности и их реализация в жизни общества и человека поставят нас перед масштабными и потенциально опасными трансформациями, которые навсегда изменят взаимодействие человека и техники, так как всякое устройство, с которым может взаимодействовать человек может обладать Общим искусственным интеллектом. Человек «всегда был и остаётся не только естественным, но и искусственным существом, то есть таким, который конституируется культурой»²³⁷, но впервые в истории своего вида человек может столкнуться с тем, что вещи, с которыми он взаимодействует, знают о нём больше, чем самые близкие люди, или даже больше, чем знает человек сам о себе. Мир станет населён не просто роботами или машинами с искусственным интеллектом, но машинами, которые обладают самым точным и интимным знанием о нас, нашей семье и наших предпочтениях. Сама эта ситуация поставит наш вид перед неизбежным переопределением того, что есть человек: человек — это то, что не может сделать машина.

²³⁶ Созанкова Е. Дипфейки: как нейросети копируют реальность // Screenlifer. URL: <https://screenlifer.com/trends/dipfejki-kak-nejroseti-kopiruyut-realnost/> (дата обращения: 16.08.2020).

²³⁷ Лекторский В. А. Человек и технонаука // Человек и культура: избр. ст. СПб.: СПбГУП, 2018. С. 339.

Заключение

Актуальность темы искусственного интеллекта и робототехники невозможно переоценить в настоящий момент. Идущая мировая гонка за лидерами в этой области — США и Китаем — подстегивает Россию и другие страны к созданию собственных передовых платформ и технологий в области искусственного интеллекта. Правильно выбранный методологический подход позволяет с одной стороны, найти возможные новые стратегические приоритеты для исследований, а с другой — не идти уже пройденным путем, сокращая затрачиваемые средства. Ведущие российские философы (В. С. Степин, В. А. Лекторский и др.) предупреждают, что эффект от искусственного интеллекта может стать «цивилизационным взрывом».

Для подготовки к такому кардинальному преобразованию нам необходима широкая общественная и научная дискуссия о назревающих вопросах и возможных ответах на них. Такую дискуссию поддерживает Научный совет при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований (НСМИИ и КИ). Совет является единственной в России организацией, которая свыше 15 лет систематически разрабатывает междисциплинарные проблемы искусственного интеллекта и когнитивных исследований. Многочисленные коллективные труды, монографии и публикации членов Совета (С. Н. Васильева, В. Л. Макарова, Д. И. Дубровского, В. А. Лекторского, А. Ю. Алексева, К. В. Анохина, В. И. Аршинова, В. Е. Лепского, О. П. Кузнецова, М. А. Холодной, В. Г. Буданова, В. Л. Васюкова) были использованы при подготовке данного диссертационного исследования.

За рубежом запущены многочисленные исследовательские программы, посвящённые философии и методологии искусственного интеллекта и робототехники. В частности, работа диссертанта защищена на ведущей мировой конференции по Общему искусственному интеллекту «AGI-2020» и

была принята к публикации в сборник работ Lecture Notes on Computer Science издательства Springer Nature²³⁸.

Основываясь на этих работах, а также на собственных разработках диссертанта, выделяются следующие значимые аспекты, которые с полным основанием (подтверждённым российскими и зарубежными публикациями, патентами) могут быть отнесены к научной новизне:

- во-первых, проведён анализ проблематики, перспективных исследований и барьеров в развитии Общего искусственного интеллекта и интеллектуальной робототехники в контексте тьюринговой методологии;

- во-вторых, впервые в научной литературе выявлены основные характеристики тьюринговой методологии, проанализированы и комплексно систематизированы частные тесты Тьюринга, выявлены основные типы для их измерения, последующих модификаций или создания новых тестов;

- в-третьих, введены новые понятия, которые дают возможность для исследователей описывать текущие и перспективные методологические подходы к искусственному интеллекту и интеллектуальной робототехнике: техно-умвелт, посттьюринговая методология, стена Тьюринга;

- в-четвёртых, рассмотрены практические и эпистемологические следствия применения введённых понятий для методологии искусственного интеллекта, которые открываются при достижении единства формы и содержания коммуникации человека и машины, а также увеличения эффективности взаимодействия человека и машины при решении совместных задач;

- в-пятых, рассмотрены потенциальные социально-культурные и мировоззренческие перспективы коммуникации людей, Общего искусственного интеллекта и интеллектуальных роботов в таких актуальных областях как производство, образование и развлечения.

²³⁸ Efimov A. Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // Artificial General Intelligence (AGI). 2020. Vol. 12177. P. 83–94. DOI: 10.1007/978-3-030-52152-3_9.

Диссертационное исследование позволяет дать обоснование посттьюринговой методологии для создания Общего искусственного интеллекта и интеллектуальной робототехники и создать методологическую основу для классификации частных тестов Тьюринга. Исследование позволяет на основе переходов из различных техно-умвельтов тьюрингового пространства по-новому подойти к разработке когнитивных архитектур для технологий интеллектуальной робототехники и Общего искусственного интеллекта.

Библиография

1. *Алексеев А. Ю.* Комплексный тест Тьюринга: философско-методологические и социо-культурные аспекты. М.: ИинтелЛ, 2013. 304 с.
2. *Алексеев А. Ю.* Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга: диссертация на соискание ученой степени доктора философских наук: 09.00.08. М., 2015. 482 с.
3. *Алексеева И. Ю., Аршинов В. И.* Информационное общество и НБИКС-революция. М.: ИФ РАН, 2016. 196 с.
4. *Анисов А. М.* Генетический метод построения теорий // *Credo new*. 2012. № 3. С. 38–53.
5. *Анохин К. В.* Когнитивные вычисления на основе нейронных гиперсетей
URL:
<https://www.youtube.com/watch?v=EXAzpc5XEiA&list=PLS18gQnIovXuRGSa-87ghTV3lNUmIH1Dk&index=12&t=0s> (дата обращения: 06.02.2020).
6. *Аронсон О. В.* Кино и философия: от текста к образу. М.: ИФ РАН, 2018. 109 с.
7. *Аршинов В. И.* Конвергентные технологии в контексте постнеклассической парадигмы сложности // *Сложность, разум, постнеклассика*. 2015. № 3. С. 42–54.
8. *Аршинов В. И.* На пути к антропному измерению NBIC-конвергенции // *Философские проблемы биологии и медицины*. Вып. 4. М., 2010 С. 41–44.
9. *Аршинов В. И.* Цифровая реальность в оптике постнеклассической парадигмы сложностности // *Проектирование будущего. Проблемы цифровой реальности: труды 1-й Международной конференции (8–9 февраля 2018 г., Москва)*. М.: ИПМ им. М. В. Келдыша, 2018. С. 147–151.
10. *Бескаравайный С. С.* Бытие техники и сингулярность. М.: РИПОЛ Классик, 2018. 476 с.

11. *Бирюков Б. В.* Кибернетика и методология науки. М.: Наука, 1964. 416 с.
12. *Бойко Е. И.* Сознание и роботы // Вопросы психологии. 1966. № 4. С. 169–177.
13. *Буданов В. Г.* Новый цифровой жизненный техноуклад — перспективы и риски трансформаций антропосферы // Философские науки. 2016. № 6. С. 47–55.
14. *Буданов В. Г., Асеева И. А.* Умвельт-анализ и дорожные карты Большого антропологического перехода // Материалы междунар. науч. конгресса «Глобалистика-2017». М.: ФГП глобальных процессов МГУ им. М. В. Ломоносова, 2017. URL: https://lomonosov-msu.ru/archive/Globalistics_2017/data/section_6_10143.htm (дата обращения: 20.06.2020).
15. *Буданов В. Г., Аршинов В. И., Лепский В. Е., Свирский Я. И.* Сложность и проблема единства знания. Вып. 1: К стратегии познания сложности. М.: ИФ РАН, 2018. 105 с.
16. *Васюков В. Л.* Искусственный интеллект и когнитивные вычисления // Актуальные проблемы современной когнитивной науки. Материалы пятой всероссийской научно-практической конференции с международным участием (18–20 октября 2012 г., Иваново). Иваново: 2012. С. 43–49.
17. *Винер Н.* Человек управляющий / Пер. с англ. Е. Панфилова, М. Аронэ, Р. Фесенко, М. Мамардашвили. СПб.: Питер, 2001. 288 с.
18. *Герасимова И. А.* Междисциплинарность и трансдисциплинарность в инженерном образовании // Философия науки и техники в России: вызовы информационных технологий / Под общ. ред. Н. А. Ястреб. Вологда, 2017. С. 48–51.
19. *Глазунов В. А.* Методологические проблемы теоретической робототехники: диссертация на соискание ученой степени доктора философских наук: 09.00.08. М., 2003. 482 с.

20. *Гоготшивили Л. А.* Непрямое говорение. М.: Языки славянских культур, 2006. 720 с.
21. *Давыдова Л. И., Конькова Г. И., Чубова А. П.* Античные мастера. Скульпторы и живописцы. Л.: Искусство, 1986. 252 с.
22. *Дамасио А. Я.* Мозг и возникновение сознания / Пер. с англ. И. Ющенко. М.: Карьера Пресс, 2018. 384 с.
23. *Деннет Д.* Насосы интуиции и другие инструменты мышления / Пер. с англ. З. Мамедьярова, Е. Фоменко. М.: АСТ, 2019. 576 с.
24. Джон Генри // Сказки народов мира. URL: <http://skazki-narodov.ru/181-dzhon-genri.html> (дата обращения: 16.08.2020).
25. *Дик Ф.* Мечтают ли андроиды об электроовцах / Пер с англ. М. Пчелинцев. М.: Fanzon, 2016. 640 с.
26. *Докинз Р.* Слепой часовщик / Пер. с англ. А. Гопко. М.: Corpus, 2015. 496 с.
27. *Долар М.* Голос и ничего больше / Пер. с англ. А. Красовец. М.: Изд-во Ивана Лимбаха, 2018. 384 с.
28. *Дубровский Д. И.* Психические явления и мозг: философский анализ проблем в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М.: Наука, 1971. 392 с.
29. *Дубровский Д. И.* Психика и мозг. Результаты и перспективы исследований // Психологический журнал. 1990. Т. 11. № 6. С. 3–15.
30. *Дубровский Д. И.* Проблема «Сознание и мозг»: теоретическое решение. М.: Канон+, 2015. 208 с.
31. *Дубровский Д. И.* Субъективная реальность // Философская антропология. 2018. Т. 4. № 2. С. 186–217.
32. *Дубровский Д. И.* Субъективная реальность и мозг. К вопросу о «Трудной проблеме сознания». URL: <https://publications.hse.ru/mirror/pubs/share/folder/bumurv2ksk/direct/79540685> (дата обращения: 21.08.2020).

33. *Дубровский Д. И.* Субъективная реальность как предмет междисциплинарного исследования // Познание и сознание в междисциплинарной перспективе. Ч. 1. М.: ИФ РАН, 2013. С. 5–27.
34. Елена. Цифровой двойник телеведущей Сбербанка. URL: https://www.tadviser.ru/index.php/Продукт:Елена_Цифровой_двойник_телеведущей_Сбербанка (дата обращения: 21.07.2020).
35. *Ефимов А. Р.* Технологические предпосылки неразличимости человека и его компьютерной имитации // Искусственные общества. 2019. Т. 14. DOI: 10.18254/S207751800007645-8. URL: <https://artsoc.jes.su/s207751800007645-8-1/> (дата обращения: 21.07.2020).
36. *Ефимов А. Р.* Посттьюринговая методология: разрушение стены на пути к общему искусственному интеллекту // Интеллект. Инновации. Инвестиции. 2020. № 2. С. 74–80. DOI: 10.25198/2077-7175-2020-2-74. URL: http://intellekt-izdanie.osu.ru/arch/2020_2_74.pdf (дата обращения: 21.07.2020).
37. *Ивлев Ю. В.* Квазифункциональность в логике и других науках // Гуманитарный вестник (МГТУ им. Н. Э. Баумана). 2020. № 6 (80). С. 1–22. URL: <https://cyberleninka.ru/article/n/kvazifunktsionalnost-v-logike-i-drugih-naukah/viewer> (дата обращения: 21.07.2020).
38. *Касавин И. Т.* Мегапроекты и глобальные проекты: Наука между утопией и технократией // Вопросы философии. 2015. № 9. С. 40–56.
39. *Канн Э.* Происхождение орудия // *Канн Э., Кунов Г., Нуаре Л., Эспинас А.* Роль орудия в развитии человека. Ленинград: Прибой, 1925. С. 21–24.
40. Кибернетика. Мышление. Жизнь / Под ред. Б. В. Бирюкова. М.: Мысль, 1964. 512 с.
41. *Киященко Л. П.* Личность как голограмма в трансдисциплинарной культуре // Вопросы философии. 2017. № 11. С. 58–68.
42. *Князева Е. Н.* Понятие «Umwelt» Якоба фон Икслюля и его значимость для современной эпистемологии // Вопросы философии. 2015. № 5. С. 30–44.

43. *Кобринский Н., Пекелис В.* Быстрее мысли. М.: Молодая Гвардия, 1963. 443 с.
44. Конференция по искусственному интеллекту. URL: <http://kremlin.ru/events/president/news/62003> (дата обращения: 20.07.2020).
45. *Крейдли Г.* Невербальная семиотика. Язык тела и естественный язык. М.: Новое литературное обозрение, 2002. 592 с.
46. Лаборатория робототехники Сбербанка. Аналитический обзор мирового рынка робототехники. URL: https://www.sberbank.ru/common/img/uploaded/pdf/sberbank_robotics_review_2019_17.07.2019_m.pdf (дата обращения 21.07.2020).
47. *Лакофф Дж.* Женщины, огонь и опасные вещи. Что категории языка говорят нам о мышлении. М.: Языки славянской культуры, 2004. 792 с.
48. *Лекторский В. А.* Совместимы ли наука и утопия? // Мифология века НТР. Утопии, мифы, надежды и реальность новейших направлений науки. От Франкенштейна и эликсира бессмертия до «биокиборгов» и постчеловека / Отв. ред. Г. А. Белкина. М.: URSS, 2019. С. 35–49.
49. *Лекторский В. А.* Философия, искусственный интеллект, когнитивные исследования // Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (г. Москва, 17–18 марта 2016 г.) / Под ред. В. А. Лекторского, Д. И. Дубровского, А. Ю. Алексева. М., 2017. С. 37–44.
50. *Лекторский В. А.* Философия. Познание. Культура. М.: Канон+, 2012. 384 с.
51. *Лекторский В. А.* Человек и технонаука // *Лекторский В. А.* Человек и культура: избр. ст. СПб.: СПбГУП, 2018. С. 331–340.
52. *Лекторский В. А.* Эпистемология классическая и неклассическая. М.: Эдиториал УРСС, 2001. 256 с.

53. *Лем С.* Сумма технологии = Summa Technologiae / Пер. с польск.; пред. В. В. Ларина; ред. и послесл. Б. В. Бирюкова и Ф. В. Широкова. М.: Мир, 1968. 608 с.
54. *Лепский В. Е.* Аналитика сборки субъектов развития. М.: Когито-Центр, 2016. 130 с.
55. *Лепский В. Е.* Асимметричный ответ информационным войнам XXI века // Рефлексивные процессы и управление. Сборник материалов XI Международного симпозиума (16–17 октября 2017 г., Москва) / Отв. ред. В. Е. Лепский. М.: Когито-Центр, 2017. С. 221–224.
56. *Лусеев И. К., Петрова Е. В., Фесенкова Л. В., Хен Ю. В.* Науки о жизни сегодня: философские инновации. М.: ИФ РАН, 2016. 239 с.
57. *Маклюэн М.* Понимание Медиа: внешние расширения человека. М.: Гиперборея, 2007. 462 с.
58. *Мамчур Е. А.* Существует ли в современном естественно-научном познании бэконовский «experimentum crucis»? // Vox. 2019. № 26. С. 141–153. URL: <https://vox-journal.org/html/issues/480/502> (дата обращения: 23.07.2020).
59. *Никифоров А. Л.* Фундаментальная наука в XXI веке // Вопросы философии. 2008. № 5. С. 58–61.
60. *Новик И. Б.* Кибернетика. Философские и социологические проблемы. М.: Госполитиздат, 1963. 208 с.
61. *Новик И. Б.* Философские вопросы моделирования психики. М.: Наука. 1969. 174 с.
62. Патент недели: может ли робот причинить вред человеку // Популярная механика. URL: <https://www.popmech.ru/technologies/487412-patent-nedeli-mozhet-li-robot-prichinit-vred-cheloveku/> (дата обращения: 10.07.2020).
63. *Парфёнов В.* ИИ победил всех в Dota2. Сразиться с ним теперь может каждый // Популярная механика. URL: <https://www.popmech.ru/technologies/news-475872-ii-pobedil-vseh-v-dota2-srazitsya-s-nim-teper-mozhet-kazhduy/> (дата обращения: 05.08.2020).

64. *Пенроуз Р.* Новый ум короля: О компьютерах, мышлении и законах физики / Пер с англ. В. О. Малышенко. М.: УРСС, ЛКИ, 2011. 400 с.
65. *Петрушенко Л. А.* Самодвижение материи в свете кибернетики. Философский очерк взаимосвязи организации и дезорганизации в природе. М.: Наука, 1971. 292 с.
66. *Поппер К.* Логика и рост научного знания. М.: Прогресс, 1983. 605 с.
67. *Рамачандран В.* Мозг рассказывает. М.: Карьера Пресс, 2005. 442 с.
68. *Розин В. М.* Понятие и современные концепции техники. М.: ИФ РАН, 2006. 255 с.
69. *Розин В. М.* Философия техники. От египетских пирамид до виртуальных реальностей: учеб. пособ. М.: NOTA BENE, 2001. 456 с.
70. *Розин В. М.* Эволюция инженерной и проектной мысли. Инженерия: Становление, Развитие, Типология. М.: Ленанд, 2015. 200 с.
71. *Риццолати Дж., Синигалья К.* Зеркала в мозге. О механизмах совместного действия и сопереживания / Пер. с англ. О. А. Куракова, М. В. Фаликман. М.: Языки славянских культур, 2012. 208 с.
72. Сайт Всемирного экономического форума в Давосе. URL: <https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2?tab=publications> (дата обращения: 28.07.2020).
73. *Сёрл Дж.* Открывая сознание заново / Пер. с англ. А. Ф. Грязнова. М.: Идея-Пресс, 2002. 256 с.
74. *Сёрл Дж.* Разумы, мозги, программы // Тест Тьюринга. Роботы. Зомби / Пер. с англ. Д. Родионова; под ред. А. Ю. Алексеева. М.: МИЭМ, 2006. 120 с.
75. *Созанкова Е.* Дипфейки: как нейросети копируют реальность // Screenlifer. URL: <https://screenlifer.com/trends/dipfejki-kak-nejroseti-kopiruyut-realnost/> (дата обращения: 16.08.2020).
76. *Стетин В. С.* Теоретическое знание. М.: Прогресс-Традиция, 2000. 744 с.
77. *Стетин В. С.* Постнеклассическая рациональность и информационное общество // Философия искусственного интеллекта. Труды

- Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (17–18 марта 2016 г., г. Москва, философский факультет МГУ им. М. В. Ломоносова). М.: ИИнтелл, 2017. С. 59–70.
78. *Стругацкий А. Н., Стругацкий Б. Н.* Понедельник начинается в субботу: Сказка для научных работников младшего возраста. М.: Детская литература, 1965. 240 с.
79. *Тавризян Г. М.* Техника, культура, человек: критический анализ концепций технического прогресса в буржуазной философии XX века. М.: Наука, 1986. 199 с.
80. *Тегмарк М.* Жизнь 3.0 / Пер. с англ. Д. Баюка. М.: Corpus, 2019. 560 с.
81. *Тищенко П. Д.* Биотехнологии инхенсмента: на пути к третьей утопии? // Рабочие тетради по биоэтике. Вып. 20: Гуманитарный анализ биотехнологических проектов «улучшения» человека: сб. науч. ст. / Под ред. Б. Г. Юдина. М.: Изд-во Московского гуманитарного университета, 2015. 168 с.
82. *Труфанова Е. О.* Информационное перенасыщение: ключевые проблемы // Философские проблемы информационных технологий и киберпространства. 2019. Т. 16. № 1. С. 4–21. URL: <https://cyberleninka.ru/article/n/informatsionnoe-perenasyschenie-klyuchevye-problemy/viewer> (дата обращения: 16.08.2020).
83. *Тьюринг А.* Может ли машина мыслить. М.: ГИФМЛ, 1960. 69 с.
84. *Тюхтин В. С.* О природе образа: психическое отражение в свете идей кибернетики. М.: Высшая школа. 1963. 121 с.
85. *Успенский В. А.* Теорема Геделя о неполноте. М.: Наука, 1982. 112 с.
86. *Финн В. К.* Искусственный интеллект: Методология, применения, философия. М.: Красанд, 2018. 418 с.
87. Фонд «Сколково», АСИ и РВК объявили о запуске конкурсов НТИ. URL: <https://upgreat.one/media-center/smi/fond-skolkovo-asi-i-rvk-obyavili-o-zapuske-konkursov-nti/> (дата обращения: 07.08.2020).

88. *Хабермас Ю.* Будущее человеческой природы / Пер. с нем. М. Л. Хорьков. М.: Весь Мир, 2002. 144 с.
89. *Харари Ю. Н.* 21 урок для 21 века / Пер. с англ. Ю. Гольдберг. М.: Синдбад, 2019. 416 с.
90. *Холодная М. А.* Естественный и искусственный интеллект. Структура и функции естественного интеллекта в контексте проблемы искусственного интеллекта // Искусственный интеллект. Междисциплинарный подход / Под ред. Д. И. Дубровского, В. А. Лекторского. М.: ИИнтелл, 2006. С. 149-163.
91. *Шалак В. И.* Социальное программирование // SocioTime / Социальное время. 2016. № 1 (5). С. 51–60.
92. *Шваб К.* Четвертая промышленная революция. М.: Эксмо, 2016. 230 с.
93. *Шульга Е. Н.* Когнитивная герменевтика. М.: ИФ РАН, 2002. 235 с.
94. *Шестакова М. А.* Методологические проблемы многоагентных систем. // Философия искусственного интеллекта. Труды Всероссийской междисциплинарной конференции, посвященной шестидесятилетию исследований искусственного интеллекта (17–18 марта 2016 г.). М.: ИИнтелл, 2017. С. 189–202.
95. *Штукельбергер К., Буданов В. Г., Олескин А. В., Колесова Л. А., Райков А. Н., Дербин Е. А., Медведев П. А., Волков А. В., Захарова Ф. Я.* Вызовы будущего: искусственный интеллект, технологии, этика // Экономические стратегии. 2019. Т. 21. № 6 (164). С. 18–29.
96. *Чалмерс Д.* Сознательный ум. В поисках фундаментальной теории / Пер. с англ. В. В. Васильева. М.: УРСС: Книжный дом «Либроком», 2013. 512 с.
97. *Черниговская Т. В.* Зеркальный мозг, концепты и язык: цена антропогенеза // Российский физиологический журнал. 2006. № 1. С. 84–99.
98. *Эверетт Д.* Как начинался язык. История величайшего изобретения. М.: Альпина Нон Фикшн, 2019. 424 с.

99. Юлина Н. С. Д. Деннет: самость как центр «нарративной гравитации» или почему возможны самостные компьютеры? // Вопросы философии. 2003. № 3. С. 104–120.
100. Юлина Н. С. Тайна сознания: альтернативные стратегии исследования // Юлина Н. С. Очерки по современной философии сознания. М.: Канон+, 2015. С. 179–221.
101. Юлина Н. С. Очерки по современной философии сознания. М.: Канон+, 2015. 408 с.
102. Ясперс К. Современная техника / Пер. М. И. Левиной // Новая технократическая волна на Западе / Сост. П. С. Гуревич. М.: Прогресс, 1986. С. 119–146.
103. Ackerman E. Can Winograd Schemas Replace Turing Test for Defining Human-Level AI? // IEEE Spectrum. URL: <https://spectrum.ieee.org/automaton/artificial-intelligence/machine-learning/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence> (дата обращения: 19.07.2020).
104. Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. 944 p.
105. Artificial Intelligence for the American People. URL: <https://www.whitehouse.gov/ai/> (дата обращения: 28.07.2020).
106. Billard A., Kragic D. Trends and challenges in robot manipulation // Science 364 (6446), eaat8414. URL: <https://science.sciencemag.org/content/364/6446/eaat8414.full?ijkey=A4elq4KH35DN.&keytype=ref&siteid=sci> (дата обращения: 28.07.2020).
107. Boden M. Grey Walter's Anticipatory Tortoises // The Rutherford journal. URL: <http://rutherfordjournal.org/article020101.html#null> (дата обращения: 16.06.2020).
108. Brooks R. A. Cambrian intelligence: The early history of the new AI. Cambridge, MA: The MIT Press, 1999. 214 p.

109. *Brooks R. A. Coherent behavior from many adaptive processes // Cliff D., Mayer J.A., Wilson S.W. Proceedings of the Third International Conference on Simulation of Adaptive Behavior, From animals to animats 3. Cambridge: MIT Press, 1994. P. 22–29.*
110. *Brooks R. A. For the Last Minute Holiday Shopper: Get an iRobot! URL: <https://corporature.wordpress.com/tag/rodney-brooks/> (дата обращения: 20.08.2020).*
111. *Brooks R. A. Intelligence Without Representation // Artificial Intelligence. 1991. No. 47. P. 139–159.*
112. *Brooks R. A. Steps Toward Super Intelligence IV, Things to Work on Now // RODNEY BROOKS Robots, AI, and other stuff. Rodney Brooks, 15 July 2018 г. URL: <https://rodneybrooks.com/forai-steps-toward-super-intelligence-iv-things-to-work-on-now/> (дата обращения: 08.08.2020).*
113. *Brooks R. A. The Case for Embodied Intelligence // Alan Turing: His Work and Impact / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 499–500.*
114. *Brooks R. A. The cog project // Journal of the Robotics Society of Japan. 1997. No. 15. P. 968–970.*
115. *Brooks R. A. Using human development as a model for adaptive robotics // Shirai Y., Hirose S. (eds.) Robotics Research. London: Springer, 1998. P. 339–343.*
116. *Brooks R. A., Breazeal C., Marjanovic M., et al. The Cog project: Building a humanoid robot // Lecture Notes in Computer Science. 1999. No. 1562. P. 52–87.*
117. *Burden D., Savin-Baden M. Virtual Humans. Today and Tomorrow. Boca Raton: CRC Press, 2019. 270 p.*
118. *Cambridge English Dictionary // Artificial Intelligence. URL: <https://dictionary.cambridge.org/ru/словарь/английский/artificial-intelligence> (дата обращения: 05.05.2020).*

119. *Carnap R.* Discussion of Critics // The Philosophy of Rudolf Carnap / Ed. by P. A. Schillp. La Salle: OpenCourt, 1963. 883 p.
120. Chinese state media's latest innovation is an AI female news anchor // CNN Business. URL: <https://qz.com/1554471/chinas-xinhua-launches-worlds-first-ai-female-news-anchor/> (дата обращения: 12.08.2020).
121. CIA Fact Book. URL: <https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html> (дата обращения: 02.06.2020).
122. *Clark A.* Can Philosophy contribute to an understanding of Artificial Intelligence? URL: <http://undercurrentphilosophy.com/medium/can-philosophy-contribute-to-an-understanding-of-artificial-intelligence/> (дата обращения: 02.08.2020).
123. *Clark A.* Reasons, robots and the extended mind // Mind & Language. 2001. Vol. 16 (2). P. 121–145.
124. *Colby K. M., Hilf F. D., Weber S.* Artificial Paranoia // Artificial Intelligence. 1971. Vol. 2. P. 1–25.
125. *Connell J.* Kitanai, Kitsui and Kiken: The Rise of Labour Migration to Japan // Economic & Regional Restructuring Research Unit. Sydney: University of Sydney, 1993. 21 p.
126. *Copeland J., Bowen J., Sprevak M., Wilson R.* The Turing Guide. Oxford: Oxford University Press, 2017. 544 p.
127. *Crafts N.* Steam as a GPT: A growth accounting perspective // The Economic journal. 2004. Vol. 114. P. 338–351.
128. DARPA Robotics Challenge. URL: https://en.wikipedia.org/wiki/DARPA_Robotics_Challenge (дата обращения: 23.07.2020).
129. *Davey P. G.* Review: «3rd symposium on theory and practice of robots and manipulators», 1980 // IEE PROC. 1980. Vol. 128. Pt. D. No. 1. P. 18.
130. *Deutsch D.* Creative Blocks. URL: <https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence> (дата обращения: 18.07.2020).

131. *Dreyfus H. L.* What computers still can't do: A critique of artificial reason. Cambridge, MA: MIT Press, 1992. 408 p.
132. *Dreyfus H. L., Dreyfus S. E.* Making a mind versus modeling the brain: Artificial intelligence back at a branchpoint // *Daedalus*. 1988. Vol. 117. No. 1. P. 15–43.
133. *Efimov A.* Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // *Artificial General Intelligence (AGI)*. 2020. Vol. 12177. P. 83–94. DOI: 10.1007/978-3-030-52152-3_9.
134. *Ekman P.* Increase your emotional awareness and detect deception. URL: <https://www.paulekman.com> (дата обращения: 18.08.2020).
135. *Emmeche C.* Does a robot have an umwelt? // *Semiotica*. 2001. No. 134. P. 653–693.
136. *Envisioning Robots in Society — Power, Politics, and Public Space* Proceedings of Robophilosophy 2018 / TRANSOR-2018 (February 14–17, 2018, University of Vienna, Austria) / Ed. by M. Coeckelbergh, J. Loh, M. Funk, J. Seibt and M. Nørskov. Amsterdam: IOS PRESS BG, 2018. 391 p.
137. *Gardels N.* USC neuroscientist Antonio Damasio argues... URL: <https://ngp.usc.edu/usc-neuroscientist-antonio-damasio-argues-that-feelings-and-emotions-are-what-make-up-human-intelligence-consciousness-and-the-capacity-for-cultural-creation-a-map-of-the-computational-mind-he-says/> (дата обращения: 12.07.2020).
138. *Gessler N.* The Computerman, The Cryptographer and The Physicist // *Alan Turing: His Work and Impact* / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 521–530.
139. *Goertzel B.* Artificial general intelligence: Concept, state of the art, and future prospects // *Journal of Artificial General Intelligence*. 2014. Vol. 5. P. 1–46.
140. *Goertzel B., Iklé M., Wigmore J.* The architecture of human-like general intelligence // *Wang P., Goertzel B.* Theoretical foundations of artificial general intelligence. Vol. 4. Amsterdam: Atlantis Press, 2012. P. 123–144.

141. *Goff P.* Galileo's Error: Foundation for a New Science of Consciousness. New York: Pantheon, 2019. 256 p.
142. *Graetz G., Michaels G.* Robots at Work. URL: http://personal.lse.ac.uk/michaels/Graetz_Michaels_Robots.pdf (дата обращения: 15.08.2020).
143. *Harnad S.* Commentary on «Computing Machinery and Intelligence» // «Parsing The Turing Test» Philosophical and Methodological Issues in the Quest for the Thinking Computer / Ed. R. Epstein, G. Roberts, G. Beber. Amsterdam: Springer Netherlands, 2008. P. 23–67.
144. *Harnad S.* Minds, Machines and Turing: The Indistinguishability of Indistinguishables // Journal of Logic, Language and Information. 2000. Vol. 9. P. 425–445.
145. *Hempell K. G.* The Logical Analysis of Psychology // *Block Ned* (ed.). Readings in Philosophy of Psychology. Cambridge: Harvard University Press, 1980. P. 1–14.
146. *Hofstadter D.* The Shallowness of Google Translate // The Atlantic, 2018. URL: <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/> (дата обращения: 26.04.2020).
147. How Google's AI Viewed the Move No Human Could Understand. URL: <https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/> (дата обращения: 29.07.2020).
148. Inside Google's Rebooted Robotics Program, The New York Times, 2019. URL: <https://www.nytimes.com/2019/03/26/technology/google-robotics-lab.html?smid=nytcore-ios-share> (дата обращения: 28.07.2020).
149. International Research Conference Robophilosophy 2020: Online conference, August 18-21, 2020. URL: <https://conferences.au.dk/robo-philosophy/> (дата обращения: 21.08.2020).
150. *Ishiguro H.* Android Science // *Kasaki M., Ishiguro H., Asada M., et al.* Cognitive Neuroscience Robotics. Tokyo: Springer Japan, 2016. P. 193–234.

151. *Ishiguro H.* Android science: Towards a new cross-disciplinary framework // Springer Tracts in advanced robotics. Berlin: Springer, 2007. P. 118–127. URL: <http://robots.stanford.edu/isrr-papers/draft/Ishiguro-final.pdf> (дата обращения: 26.08.2020).
152. *Kitano H.* Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery // AI Magazine. Spring, 2016. P. 39–49.
153. *Koch C.* The Feeling of Life Itself. Cambridge: MIT Press, 2019. 280 p.
154. *Kurzweil R.* The age of spiritual machines: When computers exceed human intelligence. New York: Penguin, 2000. 388 p.
155. Large Scale Visual Recognition Challenge (ILSVRC). URL: <http://www.image-net.org/challenges/LSVRC/> (дата обращения: 20.07.2020).
156. *Leiserson C. E., Thompson N. C., Emer J. S., Kuszmaul B. C., Lampson B. W., Sanchez D., Schardl T. B.* There's plenty of room at the Top: What will drive computer performance after Moore's law? // Science 368 (6495), eaam9744. URL: <https://science.sciencemag.org/content/368/6495/eaam9744> (дата обращения: 20.09.2020).
157. *Levesque H., Davi E., Morgenstern L.* The Winograd Schema Challenge // Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference (KR2012). Palo Alto, CA: AAAI Press, 2012. P. 552–561.
158. *Levin J.C., Maas M. M.* Roadmap to a Roadmap: How Could We Tell When AGI is a 'Manhattan Project' Away? // ECAI 2020. URL: <https://arxiv.org/abs/2008.04701> (дата обращения: 20.09.2020).
159. *Licklider J. C. R.* Man-Computer Symbiosis // IRE Transactions on Human Factors in Electronics. 1960. Vol. HFE-1. P. 4–11.
160. *Maistros G., Hayes G.* Towards an Imitation System for Learning Robots // *Vouros G.A., Panayiotopoulos T.* (eds.) Methods and Applications of Artificial Intelligence: SETN 2004. Lecture Notes in Computer Science. Vol. 3025.

- Berlin, Heidelberg: Springer, 2004. URL: https://doi.org/10.1007/978-3-540-24674-9_26 (дата обращения: 23.07.2020).
161. *Markoff J.* Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots. New York: Harper Collins Publishers, Kindle edition, 2015. 400 p.
162. *McCarthy J.* A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. URL: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (дата обращения: 26.04.2020).
163. *Menzel D. H.* The Nature of the Physical World (Eddington) // Publications of the Astronomical Society of the Pacific. 1929. Vol. 41. No. 240. P. 107–109. URL: <http://adsabs.harvard.edu/full/1929PASP...41..107M> (дата обращения: 28.07.2020).
164. *Merleau-Ponty M.* Phenomenology of perception. London: Kegan Paul, 1962. 544 p.
165. *Moravec H.* Mind Children: The Future of Robot and Human Intelligence. Cambridge, MA: Harvard University Press, 1988. 224 p.
166. *Mori M.* The uncanny valley // Energy. 1970. Vol. 7 (4). P. 33–35.
167. *Mori M., MacDorman K., Kageki N.* The uncanny valley [from the field] // IEEE Robot Automat Mag. 2012. Vol. 19 (2). P. 98–100. DOI: 10.1109/MRA.2012.2192811.
168. *Murphy R.* Introduction to AI Robotics. Cambridge: The MIT Press, 2019. 648 p.
169. *Newell A., Simon H. A.* Computer science as an empirical enquiry: Symbols and search // Communications of the ACM. 1976. Vol. 19. P. 113–126.
170. *Nikolaidis S., Lasota P. A., G. Rossano, Martinez C., Fuhlbrigge T., Shah J.* Human-Robot Collaboration in Manufacturing: Quantitative Evaluation of Predictable, Convergent Joint Action International Symposium on Robotics (ISR) 10/2013. URL: <http://interactive.mit.edu/human-robot-collaboration->

manufacturing-quantitative-evaluation-predictable-convergent-joint-action
(дата обращения: 10.07.2020).

171. NIST 2019 Speaker Recognition Evaluation. URL: <https://www.nist.gov/itl/iad/mig/nist-2019-speaker-recognition-evaluation>
(дата обращения: 20.07.2020).
172. Nöth W. Semiosis and the Umwelt of a robot // *Semiotica*. 2001. No. 134. P. 695–699.
173. O'Mahony M., Timmer M. Output. Input. And productivity measures at the industry level: the EU klems database // *The Economic Journal*. 2003. No. 119. P. 374–403.
174. Palantir exec says its work is on par with the Manhattan Project // *INPUT*. URL: <https://www.inputmag.com/tech/palantir-exec-says-its-work-is-on-par-with-the-manhattan-project> (дата обращения: 28.07.2020).
175. Penny S. What Robots Still Can't Do // *Envisioning Robots in Society — Power, Politics, and Public Space Proceedings of Robophilosophy 2018 / TRANSOR-2018 (February 14–17, 2018, University of Vienna, Austria) / Ed. by M. Coeckelbergh, J. Loh, M. Funk, J. Seibt and M. Nørskov. Amsterdam: IOS PRESS BG, 2018.* URL: https://www.researchgate.net/publication/329129638_What_Robots_Still_Can't_Do_With_Apologies_to_Hubert_Dreyfus_Or_Deconstructing_the_Technocultural_Imaginary (дата обращения: 05.08.2020).
176. Popper K., Eccles J. *The Self and Its Brain: An Argument for Interactionism*. Berlin: Springer, 1977. 597 p.
177. Putnam H. *Minds and Machines // Dimensions of Mind / Ed. by S. Hook. N.Y.: New York University Press, 1960. P. 148–179.*
178. Putnam H. The Meaning of "Meaning" // *Putnam H. Philosophical Papers. Vol. 1: Mind, Language and Reality. Cambridge: Cambridge University Press, 1975. P. 215–271.*

179. *Putnam H.* The nature of mental states // *Putnam H.* Philosophical papers. Vol. 2: Mind, language and reality. Cambridge: Cambridge University Press, 1975. P. 429–440.
180. *Richert A., Muller S., Schroder S., Jeschke S.* Anthropomorphism in social robotics: empirical results on human—robot interaction in hybrid production workplaces // *AI and Society*. 2018. No. 33. P. 412–424.
181. Robotics Roadmap and the National Robotics Initiative 2.0 (2016). URL: <https://cra.org/crn/2017/02/2016-robotics-roadmap-national-robotics-initiative-2-0/> (дата обращения: 12.07.2020).
182. *Samoili S., López Cobo M., Gómez E., De Prato G., Martínez-Plumed F., Delipetrev B.* AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. Luxembourg: Publications Office of the European Union, 2020. 90 p.
183. *Saridis G.* Intelligent robotic control // *IEEE Transactions on Automatic Control*. 1983. Vol. 28. No. 5. P. 547–557.
184. *Shneiderman B.* Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy // *International Journal of Human-Computer Interaction*. 2020. Vol. 36 (6). P. 495–504.
185. *Siciliano B., Khatib O.* Springer handbook of robotics. Switzerland: Springer International, 2016. 2160 p.
186. *Spritzer E.* Tacit Representations and Artificial Intelligence: Hidden Lessons from an Embodied Perspective on Cognition // *Müller V.C.* (ed.) *Fundamental Issues of Artificial Intelligence*. Switzerland: Springer International, 2016. P. 425–441.
187. The dawn of artificial intelligence // *The Economist*. Weekly edition. 2015. May 9. URL: <https://www.economist.com/leaders/2015/05/09/the-dawn-of-artificial-intelligence> (дата обращения: 19.07.2020).
188. The thirteenth annual conference on Artificial General Intelligence (AGI–20) take place Online, June 22–26. URL: <http://agi-conf.org/2020/call-for-papers/> (дата обращения: 29.07.2020).

189. *Thrun S., Burgard W., Fox D.* Probabilistic robotics. Cambridge, MA: MIT Press, 2005. 480 p.
190. *Turing A.* Computing machinery and intelligence // *Alan Turing: His Work and Impact* / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 562–569.
191. *Turing A.* Digital Computers Applied to Games // *Alan Turing: His Work and Impact* / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 626–644.
192. *Turing A. M.* Intelligent Machinery // *Alan Turing: His Work and Impact* / Ed. S. Cooper, J. van Leeuwen. Amsterdam: Elsevier, 2013. P. 501–517.
193. *Unhelkar V.V., Shah J.* Enabling Effective Information Sharing in Human-Robot Teams // *Conference Robotics: Science and Systems (RSS)*, 2018. URL: http://interactive.mit.edu/sites/default/files/documents/Unhelkar_RSS_Pioneers_Abstract_2018.pdf (дата обращения: 16.07.2020).
194. Use It or Lose It. URL: <https://neurosciencenews.com/brain-development-myelin-7224/> (дата обращения: 30.07.2020).
195. *Vinge V.* The Technological Singularity. URL: <https://web.archive.org/web/20100621162528/http://www.kurzweilai.net/me-me/frame.html?main=%2Farticles%2Fart0092.html> (дата обращения: 17.08.2020).
196. *Warwick K., Shah H.* Turing's Imitation Game: Conversations with The Unknow. Cambridge: Cambridge University Press, 2016. 202 p.
197. *Watt. S.* Naive Psychology and the Inverted Turing Test // *Psychology*. 1996. No. 7 (14). URL: <http://psycprints.ecs.soton.ac.uk/archive/00000506> (дата обращения: 03.08.2020).
198. Why is Toyota Developing Humanoid Robots? URL: <https://global.toyota/en/newsroom/corporate/30609642.html> (дата обращения: 26.07.2020).
199. *Wooldridge M.* The Road to Conscious Machines: The Story of AI. London: Penguin, 2020. 388 p.