

На правах рукописи

Ефимов Альберт Рувимович

**Философско-методологические основы
посттьюринговой интеллектуальной робототехники**

Специальность 09.00.08 — Философия науки и техники

АВТОРЕФЕРАТ

диссертации на соискание учёной степени

кандидата философских наук

Москва,

2020

Работа выполнена в секторе междисциплинарных проблем научно-технического развития Федерального государственного бюджетного учреждения науки Института философии Российской академии наук

Научный руководитель:

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

Официальные оппоненты:

Асеева Ирина Александровна — доктор философских наук, профессор, ведущий научный сотрудник Федерального государственного бюджетного учреждения науки Института научной информации по общественным наукам Российской академии наук.

Семёнов Алексей Львович — доктор физико-математических наук, профессор, заведующий кафедрой математической логики и теории алгоритмов Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В.Ломоносова».

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский авиационный институт (национальный исследовательский университет)».

Защита состоится «18» мая 2021 г. в «15:00» часов на заседании диссертационного совета Д.002.015.03 Института философии РАН по адресу: 109240, г. Москва, ул. Гончарная, д. 12, стр. 1, Зал заседаний Учёного совета (комн. 313).

С диссертацией можно ознакомиться в библиотеке и на сайте Института философии РАН https://iphras.ru/uplfile/zinaida/ROOTED/aspir/autoreferat/efimov/dissertatsiya_efimov_final_1_51020.pdf

Автореферат разослан «__» _____ 2021 г.

Учёный секретарь
Диссертационного совета Д 002.015.03
кандидат философских наук



Пилюгина М. А.

Общая характеристика работы

Актуальность темы исследования

Под интеллектуальной робототехникой понимают совокупность методов, способов исследований и разработки систем, интегрирующих возможности робототехники и искусственного интеллекта для создания 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, 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.

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.

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

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).

методологических подходов определяет важную роль участия философов в современных разработках искусственного интеллекта.

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) убедительно подтвердил стратегический приоритет этих технологий.

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

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

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, что напоминает эффект, сходный с изобретением паровой машины в эпоху первой промышленной революции⁵.

Стимулируя мощное развитие экономики и обещая парирование отрицательных явлений глобального кризиса, робототехника и искусственный интеллект могут быть источником новых рисков и системных проблем (нарушение неприкосновенности частной жизни, изменение самого понятия занятости и др.). Электронные средства массовых коммуникаций создают условия для неразличимости действительных фактов и лжи — искусственные нейронные сети стали основой для генерации псевдо-реальности, так называемых *deep fakes*, которые способствуют глубокому кризису экспертизы, расцвету релятивизма и скепсиса не только в обыденном знании, но и в знаниях, создаваемых в различных исследовательских центрах (В. Е. Лепский)⁶. Все это требует пристального изучения в оптике философии и методологии особенностей развития искусственного интеллекта как отрасли науки, производства и образования. Ряд аспектов, таких как

² Сайт Всемирного экономического форума в Давосе. URL: <https://intelligence.weforum.org/topics/a1Gb000000pTDREA2?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).

⁵ *Crafts N. Steam as a GPT: A growth accounting perspective // The Economic journal. 2004. No. 114. P. 338-351. DOI: 10.1111/j.1468-0297.2003.00200.x.*

⁶ *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.*

персонализация образовательных траекторий, автоматическая генерация учебных курсов и т. п. нуждаются в тщательном и всестороннем обсуждении (А. Л. Семенов)⁷.

В последнее время, несмотря на успешное развитие науки и технологий в области искусственного интеллекта, у исследователей и инженеров нет ясного ответа на вопрос, как развивать технологии в перспективе следующих десятилетий. Причина в том, что теоретико-методологическая и философская основы такого планирования разработаны 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»⁸.

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 всё это лишь частные результаты, далекие от создания Общего интеллекта.

Такое положение дел обращает нас к внимательному рассмотрению методологии, разработанной А. Тьюрингом в 1940–50-е гг., к анализу исторических ограничений этой методологии, которые должны быть преодолены на нынешнем, новом этапе развития искусственного интеллекта, когда ставится задача создания Общего искусственного интеллекта. Эти ограничения, недостатки методологии А. Тьюринга можно представить следующим образом: а) узкая интерпретация парадигмы функционализма в форме сугубо бихевиоральной и операционалистской трактовки интеллекта. С этим связаны: б) недостаточность определения критериев разумного поведения машины (тест Тьюринга); в) жесткое 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 ограничения 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, несколько не принижая при этом выдающихся эпохальных достижений А. Тьюринга).

В работах по методологии науки всегда занимало важное место осмысление парадигмы и программы новых научных исследований. Среди них при обсуждении проблематики искусственного интеллекта не раз возникали предложения опоры на так называемые «конвергентные технологии», которые включают в себя нано-технологии, биотехнологии, информационные технологии и когнитивные технологии (обозначаемые

⁷ 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).

⁸ Ясперс К. Современная техника / Пер. М. И. Левиной // Новая технократическая волна на Западе / Сост. П. С. Гуревич. М.: Прогресс, 1986. С. 119-146.

аббревиатурой NBIC), которые должны были стимулировать междисциплинарный подход к развитию искусственного интеллекта и робототехники⁹. Однако это не оправдало себя в достаточной мере (И. Ю. Алексеева, В. И. Аршинов)¹⁰, поскольку масштабные исследовательские программы фокусировались на первой части парадигмы конвергентных технологий (нано-, био-), отодвигая вторую часть (инфо-, когно-) на более далёкую перспективу.

Интеллектуальная робототехника выступает одновременно в двух качествах. С одной стороны, это область научных исследований, посвященная 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.

Искусственный интеллект наделяет любое техническое устройство возможностью изменять свое поведение на основе данных, поступающих от сенсоров, тем самым превращая его в робота: открываются новые возможности интеграции человека и робота, о которых говорил Ю. Хабермас¹¹.

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 свое существование¹². 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, идет встречный процесс, меняется не только «телесность» роботов, которая становится всё более разнообразной, в том числе и антропоморфной, но и телесность человека, которая все больше технологизируется. Это делает не эффективным использование классического рационального подхода, в рамках которого люди это только субъекты, а машины — объекты и ничего более. Исследователи искусственного интеллекта существуют в рамках саморазвивающейся системы, которой является отрасль высоких технологий, включающая сложные междисциплинарные отношения, и она создает сложные технические системы, которые предполагают в ходе функционирования многообразные кольцевые конвергентно-дивергентные связи с человеком и социальной средой. Поэтому

⁹ Аршинов В. И. На пути к антропному измерению NBIC-конвергенции // *Философские проблемы биологии и медицины*. Вып. 4. М., 2010. С. 41-44.

¹⁰ Алексеева И. Ю., Аршинов В. И. *Информационное общество и НБИКС-революция*. М.: ИФ РАН, 2016. 196 с.

¹¹ Хабермас Ю. *Будущее человеческой природы*. М.: Весь Мир, 2002. 144 с.

¹² 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.

здесь необходимо использование принципов постнеклассической рациональности, как эпистемологической основы исследования и обоснования: посттьюринговой методологии развития искусственного интеллекта и интеллектуальной робототехники.

На раннем этапе развития искусственного интеллекта и робототехники 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 существенный разрыв между теоретическими исследованиями в области философии и методологии искусственного интеллекта и практическими усилиями исследователей создавать интеллектуальные машины (роботов), которые будут обладать способностями, не уступающими человеческим. Подобный разрыв делает исследования в области методологического подхода к проблемам искусственного интеллекта и робототехники крайне актуальными и требует разработки специальных междисциплинарных программ и решения возникающих при этом острых эпистемологических вопросов (В. Г. Буданов)¹³.

Инженерно-технические проблемы (куновские «головоломки») решаются многочисленными коллективами в крупных научных центрах и предприятиях по всему миру. Но проблема долгосрочного развития Общего искусственного интеллекта и интеллектуальных роботов имеет постнеклассический междисциплинарный характер и глобальный смысл. 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 силы на создание нового поколения «умных» и полезных машин, но и реализовать потенциал искусственного интеллекта и робототехники для решения глобальных проблем нашей цивилизации.

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 операционалистского подхода к созданию «мыслящих машин» еще 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

¹³ 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.

которой имеют в виду робототехнику, способную решать сложные интеллектуальные задачи. Термин «робототехника» («robotics») появился в 1941 г. благодаря А. Азимову, хотя само слово «робот» — еще в 1920 г. у чешского писателя К. Чапека¹⁴.

С момента окончания второй мировой войны, которая дала значительный импульс развитию вычислительной техники, использовавшейся для криптоанализа и баллистических вычислений, возникла тесная связь между искусственным интеллектом и робототехникой (Д. Маккарти и Дж. Энгельбергер)¹⁵. 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 подчёркивал 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 гг. 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 МИТ 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 в. как области создания программного обеспечения, которая может помочь сделать компьютеры «интеллектуальнее», были подвергнуты значительному переосмыслению в силу разнообразных причин, включая завышенные ожидания результатов исследований в этой области (М. Мински). В 90-х гг. прошлого века стали раздаваться первые призывы к объединению направлений «искусственный интеллект» и «робототехника» в единое понятийное пространство (Р. Брукс)¹⁶.

Другой причиной появления термина «интеллектуальная робототехника» являлась постоянная критика направления, обозначаемого термином «искусственный интеллект», как якобы слишком далёкого от реальных достижений (Р. Пенроуз)¹⁷. 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

¹⁴ В пьесе К. Чапека «RUR» роботами были искусственно созданные человекоподобные биологические машины.

¹⁵ *Markoff J. Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots.* N. Y.: Harper Collins Publishers, 2015. 400 p.

¹⁶ *Brooks R. A. Intelligence Without Representation // Artificial Intelligence.* 1991. No. 47. P. 139—159.

¹⁷ *Пенроуз Р. Новый ум короля: о компьютерах, мышлении и законах физики. / Пер с англ. В. О. Малышенко. М.: УРСС, ЛКИ, 2011. 400 с.*

вполне адекватным общим термином (Р. Мёрфи)¹⁸. Таким образом, интеллектуальная робототехника является операционалистским направлением, основанным большей частью на работах Х. Моравека¹⁹, Д. Ликлидера²⁰, 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.

Следует особо отметить значительный вклад в разработку этих вопросов, внесенный советскими философами ещё в 60–80-х гг. прошлого века (Л. Б. Баженов, Б. Г. Бирюков, И. Б. Новик, Л. А. Петрушенко, В. С. Тютин, А. Г. Спиркин, А. Д. Урсул 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 учёные: П. К. Анохин, А. И. Берг, Н. А. Бернштейн, В. М. Глушков, А. Н. Колмогоров 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. Шестакова²¹), проблему эмоциональных реакций и телесности виртуальных людей (Д. Бурден, М. Савин-Баден²²), проблемы социализации роботов, их включение в публичные пространства (О. Шерер, Ж. Парвиенен²³), проблемы подходов различных культур в создании роботов, попытки осмыслить феномен «души» компьютеров (К. Джеймс²⁴). Разумеется, широко обсуждаются вопросы этики и безопасности использования роботов.

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

¹⁸ *Murphy R.* Introduction to AI Robotics. Cambridge: The MIT Press, 2019. 648 p.

¹⁹ *Moravec H.* Mind Children: The Future of Robot and Human Intelligence. Cambridge, MA: Harvard University Press. 1988. 224 p.

²⁰ *Licklider J.C.R.* Man-Computer Symbiosis // IRE Transactions on Human Factors in Electronics. 1960. Vol. HFE-1. P. 4-11.

²¹ *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.

²² *Burden D., Savin-Baden M.* Virtual Humans. Today and Tomorrow. Boca Raton: CRC Press, 2019. 270 p.

²³ *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.

²⁴ *James K.* The Battle for the Robot Soul // Philosophy Now. 2020. No. 139. P. 16–19.

представлена в нашей философской литературе, хотя она является важным условием её дальнейшего эффективного развития, особенно когда речь идет о прорывных достижениях в этой области.

Важным стимулом для формирования посттьюринговой методологии интеллектуальной робототехники явились труды А. Ю. Алексеева²⁵ в области комплексного теста Тьюринга.

Диссертация опирается на результаты исследований в области философских проблем искусственного интеллекта и смежных с ними вопросов (особенно эпистемологического характера), поставленных 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 диссертации 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 в области 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.

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

²⁵ 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.

методологии А. 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.

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

Научная новизна диссертационного исследования может быть сформулирована следующим образом:

- проведен 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;
- впервые 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 диапазон возможностей робота («мир робота», т. е. тот срез отображения мира, каким его «видит» интеллектуальный робот, а также набор действий, доступных для него в этом «мире»). При этом выделены и описаны четыре «техно-умвельта», которые охватывают все мыслимые диапазоны действий робота;
- предложен новый подход для развития научных исследований в области интеллектуальной робототехники на основе посттьюринговой 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.

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

Источники диссертации

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

Теоретической и методологической основой исследования выступает, в первую очередь, развитая в российской философии концепция постнеклассической эпистемологии в её отношениях 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.

Положения, выносимые на защиту:

- 1 Введен 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 создаёт возможности для рассмотрения методологии Тьюринга в оптике постнеклассической научной рациональности и выяснения её исторической ограниченности.
- 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.

Теоретическая и практическая значимость исследования. Выполненное в данной диссертационной работе исследование позволило теоретически обосновать посттьюринговую 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, опубликованные статьи и полученный автором патент подтверждают значение выполненного исследования, так как служат расширению деятельных возможностей «воплощённого интеллекта» и его способностей невербального общения в физическом и виртуальном мире при взаимодействии с человеком.

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 и практические результаты;
- согласованностью выводов с существующими актуальными фундаментальными исследованиями в области искусственного интеллекта, робототехники, постнеклассической эпистемологии и синергетики;
- апробацией результатов исследования.

Апробация результатов исследования

Диссертация была обсуждена и рекомендована к защите в секторе междисциплинарных проблем научно-технического развития Института философии РАН.

Результаты диссертационного исследования обсуждались на заседаниях сектора междисциплинарных проблем научно-технического развития Института философии РАН и были представлены в форме докладов на международных и всероссийских конференциях:

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

Структура и объём работы. Диссертация состоит из введения, трех глав, заключения и библиографии, объём составляет 165 стр. Список литературы включает в себя 199 наименований источников.

Основное содержание работы

Во «Введении» обосновывается высокая актуальность темы исследования, раскрывается степень научной разработанности проблемы, определяются цель и задачи исследования, теоретические и методологические основы авторского анализа, научная новизна, излагаются основные положения, выносимые на защиту, отмечаются теоретическая и практическая значимость работы, её апробация.

Глава 1. «Методологический анализ современных исследований в интеллектуальной робототехнике» состоит из трёх параграфов.

В параграфе 1. «Обобщённый обзор современной интеллектуальной робототехники» дается общий аналитико-методологический обзор интеллектуальной робототехники и особенностей её применения для решения актуальных задач современного общества и социальной деятельности человека. Рассматривается операционалистское определение интеллектуальной робототехники, согласно которому любое техническое изделие человека является интеллектуальным роботом, если обладает тремя способностями одновременно: 1) способностью воспринимать окружающий мир с помощью сенсорных приспособлений (видеокамеры, сонары, лазерные дальномеры, радары 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 путем манипуляции объектами физического мира и/или собственного перемещения. Анализируется ряд примеров использования робототехники как с точки зрения практического применения, так и с точки зрения теоретических основ исследований этого направления (Р. Мёрфи, С. Трун²⁶, Б. Сицилиано, О. Хаттиб²⁷).

В параграфе 2. «Основные тренды современной интеллектуальной робототехники в свете методологических проблем создания Общего искусственного интеллекта» проводится анализ основных трендов в разработке Общего искусственного интеллекта и разбор методологических проблем, связанных с использованием робототехники как инструмента исследования и конструирования новых подходов в этом направлении. Обосновывается важность исследований в области искусственного интеллекта как государственной задачи, закреплённой в национальных программах ряда государств, включая Российскую Федерацию. Поддержка исследований в области искусственного интеллекта имеет высший приоритет важности в России, заданный Президентом В. В. Путиным, и является по сути «мегапроектом» (И. Т. Касавин) для России²⁸. Дается рабочее определение интеллекта, как «интегративной способности упорядочивать чувственные данные, обоснованно рассуждать и нормативно регулировать поведение» (А. Ю. Алексеев). Расширением этого определения является модель В. К. Финна, которая показывает, что ряд способностей естественного интеллекта может быть осуществлен на небиологической основе в полностью автоматическом режиме без участия человека, тогда как некоторые способности (такие как целеполагание,

²⁶ Thrun S., Burgard W., Fox D. Probabilistic robotics. Cambridge, MA: MIT Press, 2005. 480 p.

²⁷ Siciliano B., Khatib O. Springer handbook of robotics. Springer, 2016. 2160 p.

²⁸ Касавин И. Т. Мегапроекты и глобальные проекты: Наука между утопией и технократией // Вопросы философии. 2015. № 9. С. 40–56.

любопытство и др.) могут выполняться лишь в интерактивном режиме с участием человека ²⁹. Даются описания технологических трендов, которые оказывают фундаментальное влияние на исследования Общего искусственного интеллекта и робототехники. В частности, появление дешёвых систем технического зрения значительно упростило создание новых типов роботов, поведение которых основано на визуальной оценке окружающей среды и адаптации моделей поведения, основанных на получаемой информации. Кроме того, имеющиеся наборы данных (массивы текстовой, визуальной информации) размечаются для дальнейшего обучения нейронными сетями в промышленных масштабах. Это делает доступным использование многослойных искусственных нейронных сетей. Получили широкое распространение ускорители вычислений, основанные на математическом обеспечении, применяемом для расчётов в компьютерной графике. Важно подчеркнуть, что когнитивные науки достигли значительных успехов в изучении механизмов работы мозга человека. Ряд теорий, описывающие работу сознания и мозга человека, получают достаточное экспериментальное подтверждение (Дж. Риццоллати, А. Дамасио, К. Кох, В. Рамачандран и др.). В частности, мозг человека несёт в себе кодовые структуры множественных образов и действий в окружающем мире. Их расшифровка с помощью нейронаучных методов открывает новые возможности моделирования сложных интеллектуальных функций в целях развития интеллектуальной робототехники и разработки Общего искусственного интеллекта (Д. И. Дубровский) ³⁰. На основе обзора применения интеллектуальной робототехники во время пандемии COVID-19 делается вывод о возросшей актуальности автономности и надёжности искусственного интеллекта и роботов (А. Р. Ефимов)³¹.

В параграфе 3. «Теоретические и эпистемологические трудности развития интеллектуальной робототехники и Общего искусственного интеллекта» среди теоретических трудностей в первую очередь отмечается эпистемологическая проблема субъектности искусственного интеллекта и, соответственно, продуктов робототехники. 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 реализованных моделей искусственного интеллекта. Обсуждение этой проблематики началось еще с А. Тьюринга, который заменил вопрос «может ли машина мыслить» на операциональный тест того, что принято считать мыслительным актом. Эта установка положила начало философскому направлению функционализма и имела принципиальное методологическое значение для всего хода развития искусственного интеллекта и многих отраслей научного знания. В аналитической философии это направление получило развитие в трудах Х. Патнэма, Дж. Фодора и др. Однако оно было подвергнуто критическому анализу в работах Дж. Сёрла и Т. Нагеля, которые критиковали машинный функционализм за игнорирование эпистемологических вопросов ценностно-смыслового плана (подчеркивая отсутствие понимания компьютером смысла воспринимаемой

²⁹ Финн В. К. Искусственный интеллект: Методология, применения, философия. М.: Красанд, 2018. С. 37–38.

³⁰ Дубровский Д. И. Проблема «Сознание и мозг»: Теоретическое решение. М.: Канон+, 2015. 208 с.

³¹ 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.

информации), за крайне упрощенную трактовку понятия естественного интеллекта. Этими философами и их последователями были высказаны принципиальные возражения против создания машинного интеллекта, имеющего возможности, сопоставимые с мышлением человека.

В отличие от концепций представителей аналитической философии, которые обсуждались выше, советские философы ещё в 60–70 гг. прошлого столетия 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. Тьютин³², Б. В. Бирюков³³, И. Б. Новик³⁴, Л. А. Петрушенко³⁵, Д. И. Дубровский³⁶ и мн. др.³⁷), специально рассматривали вопросы робототехники (Е. И. Бойко³⁸).

Большие сомнения в возможности создания сильного искусственного интеллекта получили эмпирическую поддержку в 70-х годах прошлого века в связи с описанием так называемого эффекта «зловещей долины», согласно которому робот, имеющий значительное сходство с человеком, но далёкий от тождественности с ним, естественным образом отторгается нашей психикой как субъект и не заслуживает такого же доверия, как при коммуникации с человеком (М. Мори, К. МакДорман, В. Хэнсон и др.). Трудности создания не только естественного, но и Общего искусственного интеллекта (и соответствующего им типа роботов) многократно обсуждались в философской и специальной литературе, не достигая общепринятого решения.

В этой связи автором рассматриваются вопросы использования принципов применимости постнеклассической рациональности для осмысления взаимодействия робота и человека (В. С. Степин, В. Г. Буданов).

Глава 2. «Роль тьюринговой методологии в развитии интеллектуальной робототехники» посвящена рассмотрению функционалистских принципов тьюринговой методологии — её возникновению, роли в развитии искусственного интеллекта и интеллектуальной робототехники.

В параграфе 1. «Определение тьюринговой методологии в контексте интеллектуальной робототехники» рассматривается операционалистский подход А. Тьюринга к вопросу «могут ли машины мыслить?» на основе идей, которые им были

³² Тьютин В. С. О природе образа: психическое отражение в свете идей кибернетики. М.: Высшая школа, 1963. 121 с.

³³ Бирюков Б. В. Кибернетика и методология науки. М.: Наука, 1964. 416 с.

³⁴ Новик И. Б. Кибернетика. Философские и социологические проблемы. М., 1963. 208 с.; Новик И. Б. Философские вопросы моделирования психики. М.: Наука, 1969. 174 с.

³⁵ Петрушенко Л. А. Самодвижение материи в свете кибернетики. Философский очерк взаимосвязи организации и дезорганизации в природе. М.: Наука, 1971. 292 с.

³⁶ 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.

³⁷ Кибернетика. Мышление. Жизнь / Под ред. Б. В. Бирюкова. М.: Мысль, 1964. 512 с.

³⁸ Бойко Е. И. Сознание и роботы // Вопросы психологии. 1966. № 4. С. 169–177.

высказаны в фундаментальной статье в журнале *Mind* в октябре 1950 г., а также в более ранних и менее известных публикациях. А. Тьюринг не исключал создания воплощённых физически интеллектуальных машин, но не видел технической возможности для их реализации.

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 (Д. Денет и др.). Такого рода интерпретация о «ненужности» сознания, близкая к бихевиористским установкам, была присуща также А. Тьюрингу. На первом этапе развития искусственного интеллекта она не тормозила исследований: задачи описываются в понятиях функциональных отношений, 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 рассмотрение феноменов сознания не является необходимым элементом. Однако создание Общего (универсального) искусственного интеллекта и робототехники, которая основана на его использовании, требует иных подходов и новых методологических решений.

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», сформулированных А. Тьюрингом уже в начале 1950-х (игры, криптография, изучение языков, переводы, решение математических задач и вербальное взаимодействие) существенно расширился. В результате применения этого подхода возникло большое семейство частных тестов Тьюринга, которые пытаются ответить на частный вопрос: «может ли машина выполнить определённое действие?». Такого рода частные тесты Тьюринга слабо упорядочены, не имеют единой методологии и нуждаются в основательной классификации. Эта задача составляет важное условие разработки посттьюринговой методологии.

В параграфе 2. «Методологическая роль классификации частных тестов Тьюринга в развитии интеллектуальной робототехники» автором впервые в отечественной литературе предложена классификация частных тестов Тьюринга в соответствии с подходом, основанным на идее имманентных для машин пространств. Диссертантом предложено соотнести всё множество частных тестов Тьюринга с понятием *Umwelt* (Я. фон Икскуль, Е. Н. Князева, В. Г. Буданов)³⁹ по единой методологии. Вводится

³⁹ Термин заимствован из этологии (Я. фон Икскуль). См.: Князева Е. Н. Понятие «Umwelt» Якоба фон Икскуля и его значимость для современной эпистемологии // Вопросы философии. 2015. № 5. С. 30–44; Буданов В. Г., Асеева И. А. Умвельт-анализ и дорожные карты Большого

понятие *техно-умвельт*, которое определяет «мир робота» в тьюринговом пространстве, т. е. содержит тот срез реального отображения мира, каким его «видит» робот, искусственный интеллект (или машина), а также набор действий, доступных для данного интеллектуального робота (или машины). Согласно предложенному подходу, все частные тесты Тьюринга классифицируются по четырём классам техно-умвельтов, расположенным вдоль двух фундаментальных осей: *виртуальное-физическое* и *невербальное-вербальное*. Таким образом, можно показать, что все созданные ранее или создаваемые в будущем частные тесты Тьюринга классифицируются по четырём техно-умвельтам, образующим тьюринговое пространство человеко-машинного взаимодействия: 1) *виртуальное-вербальное*; 2) *виртуальное-невербальное*; 3) *физическое-невербальное*; 4) *физическое-вербальное*. Каждое из этих пространств характеризуется срезом восприятия и множеством доступных действий, которые проявляются в частном тесте Тьюринга, который проходит робот, способный к действиям в данном техно-умвельте. Показано, что мир, в котором оперирует робот данного техно-умвельта является ограниченным для него изначально. Тьюринговая методология основывалась на бихевиоральном подходе, подразумевавшим исключение сознания и наблюдателя: важно не то, какие машина даёт ответы (или действия), но важно лишь то, что она даёт схожие с человеком ответы (действия) для аналогичной ситуации. Однако в этом проявляется фундаментальный барьер, который необходимо преодолеть для выхода из ограничений парадигмы машинного функционализма, ставшего основой тьюринговой методологии. Преодоление недостатков тьюринговой методологии может стать основой для достижения значимого прогресса в области искусственного интеллекта и робототехники.

В параграфе 3 «Стена Тьюринга как фундаментальный философско-методологический барьер развития интеллектуальной робототехники» проводится детальный разбор тьюринговой методологии и конструкции, основанной на тесте «может ли машина мыслить?», предложенном им в 1950 г. В ходе анализа автор показывает существенность элемента «стена Тьюринга», ранее ускользавшего от внимания исследователей. Стена Тьюринга, разделяющая субъект испытания (судья) и объект испытания (человек или машина) является фундаментальным эпистемологическим барьером на пути развития искусственного интеллекта и интеллектуальной робототехники. Причина этого в том, что стена заставляет робота (и опосредованно создателей роботов или искусственного интеллекта) реализовывать только те задачи, которые могут быть решены роботом в полном отрыве от человека. Однако современные исследования в различных областях человеко-машинного взаимодействия (В. Унхелькар, Дж. Ша и др.)⁴⁰ говорят о том, что максимальный эффект от использования роботов или искусственного интеллекта достигается при совместном использовании, т. е. в коллаборации человека и машины. Разделение человека и машины при операционалистском сравнении является ненужным ограничением, которое тормозит развитие всей области искусственного интеллекта и

антропологического перехода // Материалы международного научного 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).

⁴⁰ *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).

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.

Прослеживается и раскрывается аналогия между постнеклассической триадой субъект — средство наблюдения — объект и посттьюринговой методологией человек — стена Тьюринга и тест Тьюринга — робот, в которой фокус исследования сдвинут на субъектность взаимодействия человека с роботом и обратную связь этого взаимодействия, минуя стену Тьюринга.

Глава 3. «Перспективы применения посттьюринговой методологии в интеллектуальной робототехнике» состоит из трёх параграфов.

В параграфе 1. «Посттьюринговая методология как концептуальный инструмент разрушения стены Тьюринга» обсуждается роль и структура новой методологии как необходимое условие дальнейшего прогресса в разработке Общего искусственного интеллекта. Концепция воплощённого искусственного интеллекта находит свое отражение во многих работах современных исследователей (Р. Брукс, Б. Герцль и др.). Однако ранее все предложения исследователей концентрировались лишь на одном избранном «техно-умвельте», который зачастую отражал лишь специализацию авторов в том или ином виде его деятельности. Предложенный диссертантом новый подход заключается в создании таких когнитивных архитектур и воплощении их в таких механизмах, которые сразу сфокусированы на нескольких «техно-умвельтах» — роботы (или в общем случае искусственный интеллект) могут 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 вследствие универсализма разрабатываемых систем. Он требует более широкого использования для создания соответствующих когнитивных архитектур результатов современных феноменологических исследований динамической структуры явлений сознания (Д. И. Дубровский)⁴¹. В определенной степени такой подход отражает процесс формирования психики в ходе биологической эволюции. Теоретически допустимо, что система, способная воспринимать мир и действовать во всех четырёх типах «техно-умвельтов», может обладать собственным сознанием. Система, действующая лишь в одном типе техно-умвельта, в принципе не способна обладать сознанием, так как не может абстрагировать феномены и самостоятельно создавать понятия.

На современном этапе развития науки использование результатов феноменологии сознания и достижений в области нейронаучных исследований является необходимым

⁴¹ Дубровский Д. И. Субъективная реальность // Философская антропология. 2018. Т. 4. No. 2. С. 186–217. DOI: 10.21146/2414-3715-2018-4-2-186-217. URL: <https://pa.iph.ras.ru/article/view/3348/2430> (дата обращения: 18.07.2020).

условием построения новых посттьюринговых когнитивных архитектур для создания интеллектуальных роботов и Общего искусственного интеллекта.

В параграфе 2. «Экспериментальный интеллектуальный робот «Э.ЛЕНА» как демонстрация применимости посттьюринговой методологии» рассматривается применение посттьюринговой методологии при создании экспериментального интеллектуального робота-телеведущего «Э.ЛЕНА». Как было показано выше, А. Тьюринг предлагал создавать интеллектуальные машины, работающие лишь в одном типе умвельта **«виртуальное-вербальное»**, операционное пространство которых ограничено лишь приёмом и передачей символической вербальной информации. Инженерному коллективу под руководством диссертанта удалось создать и запатентовать экспериментального робота, который сделал возможным объединение двух «техно-умвельтов» — традиционного для теста Тьюринга **«виртуального-вербального»** и нового «техно-умвельта» — **«виртуального-невербального»**, ранее использовавшегося лишь в компьютерных играх. Робот «Э.ЛЕНА» является виртуальным гуманоидом, который обладает виртуальным гуманоидным телом, наделённым мимикой, может использовать естественный язык, полностью поддерживая русский язык (виземы, фонемы), является автономным (не требующим действий оператора при выполнении базовых функций), обладающим собственной персоной (в зависимости от персоны чат-бота) и имеющим первичную возможность для обучения изменениям внешнего мира (в роботе реализовано распознавание лиц собеседников робота). Впервые в научной литературе представлена архитектура подобного робота. Уникальность её заключается в том, что при её построении использовался не только ряд существенных характеристик субъективной реальности человека (опирающихся на современные исследования феноменологии сознания), но и сложный комплекс внешних проявлений сознательной деятельности человека, исполняющего роль ведущего на телевидении: мимика, выражение глаз, жесты, интонации голоса и т. д.

Опорой для этого служили исследования по проблеме невербальных коммуникаций Г. Е. Крейдлина, классические идеи М. М. Бахтина, основательные философские исследования проблемы «Непрямого говорения» Л. А. Гоготишвили, ряд работ, посвящённых, философскому и психологическому осмыслению медиа-дискурса. Опыт создания «Э.ЛЕНЫ» показал важность использования результатов философских, психологических, лингвистических исследований сознания и языка для обоснования посттьюринговой методологии и эффективного продвижения в сторону Общего интеллекта.

В параграфе 3. «Мировоззренческие перспективы посттьюринговой коммуникации людей и роботов» высказываются гипотетические соображения о будущем взаимодействии роботов и людей, исходя из предположения о технологической неразличимости роботов от людей. Подобные сюжеты давно и широко обсуждаются в футурологической и философской литературе (В. Г. Буданов и др.)⁴². Здесь обычно сочетаются разноплановые вопросы. Одно дело, когда речь идет о неразличимости человеком-наблюдателем поведения человека и робота в некоторых конкретных случаях, другое — когда во всех случаях, т. е. речь тогда идет фактически о тождественности человека и робота. Но это означает, что в ходе антропотехнологической эволюции возник

⁴² Буданов В. Г. Новый цифровой жизненный техноуклад — перспективы и риски трансформаций антропосферы // Философские науки. 2016. № 6. С. 47–55.

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 всё 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 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-умельцах 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 интеллекта.

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

В заключении подводятся итоги диссертационного исследования под углом связи проблем сознания, Общего искусственного интеллекта, интеллектуальной робототехники с развитием посттьюринговой методологии. Отмечается, что на основании разработанного концептуального подхода возможно достижение существенных результатов: в построении новых гибридных когнитивных архитектур для Общего искусственного интеллекта и интеллектуальной робототехники; в разработке эффективных операционных тестов для определения прогресса в создании новых видов роботов и технологий искусственного интеллекта; а также в более основательном осмыслении грядущих коммуникаций между человеком и роботом и вопросов этического подхода к созданию Общего искусственного интеллекта.

Основные положения диссертации отражены в следующих публикациях автора:

Статьи в периодических изданиях, включенных в Перечень рецензируемых научных изданий, рекомендуемых ВАК:

1. *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).
2. *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/>.
3. *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.

Статьи, входящие в международные реферативные базы данных и системы цитирования (Scopus и Web of Science):

1. *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.

Патенты Российской Федерации на изобретения:

1. *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);
2. *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. Затыговым).

Публикации в других научных изданиях:

1. *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.
2. *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>.
3. *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.
4. *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.
5. *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>.
6. *Ефимов А. Р.* Боевые роботы: угрозы учтённые или непредвиденные // Индекс безопасности. 2016. Т. 22. № 3–4. С. 79–96.