

ОТЗЫВ
на диссертацию
Ефимова Альберта Рувимовича
«Философско-методологические основы посттьюринговой
интеллектуальной робототехники»,
представленную на соискание ученой степени кандидата философских наук по специальности 09.00.08. – философия науки и техники.

Диссертация соответствует указанной специальности и соответствующей отрасли науки.

Актуальность исследования определяется очередным взрывным ростом интереса к проблеме искусственного интеллекта и его применениям, что характеризуется взрывным ростом числа публикаций и патентов по данной тематике, а также увеличением объема инвестиций в данную область и числом государственных программ инициированных различными государствами — за последние несколько лет. Одной из важнейших задач, имеющих в повестке дня — создание так называемого сильного искусственного интеллекта или общего искусственного интеллекта, в также — интеллекта человеческого уровня. При этом одной из составляющих данной задачи является определение рамок и критериев, определяющих самое решение данной задачи. С практической точки зрения, важнейшей областью приложения систем искусственного интеллекта является робототехника. Рассматриваемая диссертация непосредственно посвящена разбору подходов и критериев машинного интеллекта и верификации его достижения с различных позиций, применительно к интеллектуальной робототехнике. Диссертация впервые в отечественной литературе всесторонне разбирает проблему как самого теста Тьюринга, и связанные этим аспекты, так и решения, позволяющие расширить границы и поднять рамку при определении возможностей интеллектуальных робото-технических систем.

В связи со сказанным выше, диссертация Ефимова Альберта Рувимовича представляется весьма востребованной, актуальной и оригинальной.

Диссертация состоит из введения, трех глав, каждая из которых состоит из трех частей, а также, заключением и списком литературы.

Первая глава диссертации посвящена обзору предыдущих исследований и анализу современных подходов к в интеллектуальной робототехнике и уточнению определений. Заслуживает внимания богатый библиографический обзор актуальной литературы по данной тематике. Автор справедливо указывает на множественность подходов к

определению искусственного интеллекта и сознания и разбирает позиции отдельных авторов с целью создания методологической платформы для дальнейших исследований.

Во второй главе автор предлагает разделить понятие «тьюринговой» (классической) и «пост-тьюринговой» (перспективной) методологий верификации достижения искусственного интеллекта. Последняя представляется как оригинальная разработка самого автора. Для её обоснования автор разбирает и классифицирует различные типов тьюринговых тестов, обсуждавшихся с середины прошлого века. Данная Эта классификация приводит автора к понятию «техноумвельта», предлагаемое в качестве обобщенного высокоуровневого определения «среды», в которой предполагается как функционирование интеллектуального робота так и решение вопроса о степени его интеллектуальности. Далее автор предлагает в качестве основы систему из четырех фундаментальных «техноумвельтов», позволяющих определять интеллектуальность роботов в вербальных, невербальных, виртуальных и физических средах.

В третьей главе автор развивает свою методологию построения интеллектуальных роботов на основе их «воплощения» в системе базовых «техноумвельтов» определенных ранее и обсуждает её апробацию на мульти-модальном вербально-невербальном роботе собственной разработки. Далее, выводя возможность появления сознания у роботов, способных к операционной деятельности во всех указанных средах, автор обсуждает практические аспекты внедрения подобных роботов в повседневную жизнь.

Диссертация представляет собой оригинальное исследование на переднем крае философской мысли и методологических разработок, возвращая научно-практическое сообщество к необходимости пересмотра «теста Тьюринга», как то было сделано несколько десятилетий назад в концепции «детский тест Тьюринга» (В.Парте). Вместе с тем, в отличии от тезиса В.Парте о необходимости обучения как элемента теста, автор делает упор на расширенном понятии «мульти-модальности», задаваемой на базе указанных «техноумвельтов».

Научная новизна диссертации в том, автором дает обоснование методологии разработки и тестирования искусственного интеллекта и его приложений в робототехнике, что может создать основу для новых технологических решений как в области построения общего и сильного искусственного интеллекта как такового, так и для создания прикладных решений в интеллектуальной робототехнике.

Диссертация представляет собой теоретически и практически значимое актуальное исследование, обладающие необходимым научной новизны. Диссертантом опубликовано

несколько статей, включая статью в сборнике посвященном общему искусственному интеллекту, с представлением статьи на международной конференции AGI-2021 по данной тематике. Авторские идеи, изложенные в диссертации, включены в монографию «Сильный искусственный интеллект: на подступах к сверхразуму» («Альпина Паблишер», 2021).

Часть положений автора побуждает к дальнейшей дискуссии. Отмечу три вопроса, которые хотелось бы принять во внимание в дальнейших исследованиях диссертанта:

1. При всей обоснованной значимости для систем сильного или общего искусственного интеллекта «мульти-модальности» или способности к операционной деятельности в различных «техноумweltах», не упускается ли требование «детского теста Тьюринга» где на первое место выдвигается способность к обучению? Впрочем, допускаю встречную аргументацию автора, что верифицировать абстрактную способность к обучению методологически сложнее, чем создать набор базовых «сред», функционирующую в которых испытываемая система должна иметь возможность обучиться.

2. Не является ли выделение вербальных «умweltов» данью антропоморфного подхода к проблеме искусственного интеллекта, где его носитель должен непременно быть способен к вербальному взаимодействию с другими агентами в его окружении, то есть - в социальной среде? Репрезентативный ответ на данный вопрос, однако, скорее всего, потребовал бы отдельного исследования как социальной функции и обусловленности интеллекта вообще, так и обсуждения связи между социальной, семантической и лингвистической его составляющими, что не позволило бы ограничиться рамками данного исследования.

3. Является ли разделение на «виртуальные» и «физические» «умweltы» принципиальным с точки зрения принципиальной проблемы искусственного интеллекта, где разнообразие «сред» определяется набором параметров и сложностью законов их изменения? Если разнообразие сред и их сложность первичны, то, при достаточно высокой точности построения «виртуальной» модели некой физической среды, возможности агента осуществлять интеллектуальную деятельность в «виртуальной» среде будут примерно эквиваленты возможностям действия в её «физическом» варианте. Тогда более важным будет разнообразие сред как таковых, с точки зрения их параметрических моделей, а не то, являются ли они «виртуальными» или «физическими».

Вместе с тем, заданные выше вопросы не умаляют значимости работы, а только лишь подчеркивают сложность решаемой проблемы и неоднозначность решения ряда

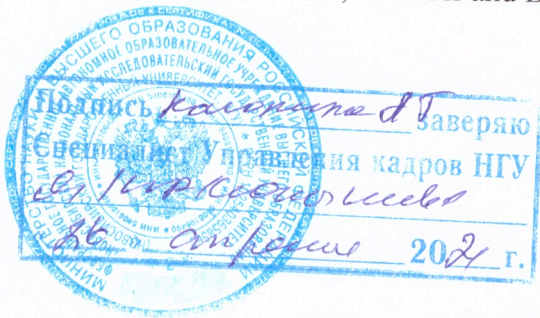
затрагиваемых ею проблем, и потому не снижают высокую оценку исследования и его результаты.

Выполненная работа представляет научную и практическую значимость и отвечает требованиям, установленным Министерством науки и образования Российской Федерации. Проведенное исследование говорит о глубоком понимании автором темы и его высокой квалификации. По своему содержанию, структуре, завершенности, обоснованию теоретических положений и практического использования результатов полностью соответствует требованиям п.п.9 и 10 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства Российской Федерации от 23 сентября 2013 г. № 842, а её автор – Ефимов Альберт Рувимович – заслуживает присуждения учёной степени кандидата философских наук по специальности 09.00.08 – философия науки и техники.

Кандидат технических наук,

Колонин Антон Германович, к.т.н.

Новосибирский Государственный Университет, ведущий специалист
SingularityNET Foundation, Lead AI and Blockchain Architect



Отзыв на диссертацию на соискание учёной степени кандидата философских наук «**Философско-методологические основы пост-тьюринговой интеллектуальной робототехники**», подготовленную Альбертом Рувимовичем Ефимовым

Представленное диссертационное исследование является предельно актуальным, поскольку, посвящено проблеме взаимодействия искусственного интеллекта (и в том числе, специально сконструированных физических объектов, управляемых искусственным интеллектом – «роботов») с человеком как носителем естественного интеллекта и сознательной воли. И социально-экономические, и социокультурные тенденции последней четверти XX века – первой четверти XXI века говорят о том, что вероятность достижения человеком оптимального результата своих действий всё в большей степени обуславливается использованием искусственного интеллекта, представленного в разнообразных инструментальных формах: от робота на производстве до программы, позволяющей в максимально короткое время дать ответ на проблемную задачу на основе громадного массива данных (Big Data). При этом, философская мысль до сих пор не дала отчётливого ответа на вопрос как о пределах возможностей искусственного интеллекта, так и о его действительном соотношении с базисным, человеческим интеллектом (что породило как популярные в массовой культуре образы «восстания машин», так и вполне реальное снижение темпов в разработке конкретных технологических систем, основанных на искусственном интеллекте).

Отметим, что с точки зрения представляемой нами философии самоопределения, и в целом с точки зрения философской антропологии, вопрос о соотношении возможностей «естественного» (человеческого) и «искусственного» (машинного) интеллекта принципиально важен. Сама по себе разработка этой проблемы позволяет более точно определить те базовые характеристики, которые характеризуют человека, сохраняются в отсутствие любых сопутствующих человеку феноменов и систем, в том числе, созданных им самим, - и именно такие характеристики могут и должны, на наш взгляд, рассматриваться как базис и основание для самоопределения человека. С другой стороны, в ситуации, когда фактическими «функциональными органами» человека (если использовать терминологию культурно-исторического подхода в психологии) стали машины, управляющие ими программные разработки, системы виртуальной реальности, человек в процессе своего самоопределения начинает воспринимать все эти технические объекты не только как инструменты, но и, зачастую, как компоненты своего само-образа, статус которых в рамках этого самообраза требует уточнения.

Представленное исследование А.Р. Ефимова позволяет существенно продвинуться в решении поставленных выше проблем. Он заведомо ставит вопрос о возможностях (пределе возможностей) искусственного интеллекта и управляемых им устройств, и далее, о соотношении их с «изначальным», «естественным» интеллектом, двояко: с инструментально-технологической и с антропологической точки зрения. В качестве основной проблемы он рассматривает применимость в наши дни известного «теста Тьюринга», позволяющего оценить равенство или неравенство «интеллектуальных» возможностей человека и программного продукта (либо программно управляемого устройства) за счёт серии формализованных процедур. Именно А. Тьюринг, как показал автор, с одной стороны, доказывал принципиальную возможность равенства человека и машины в их интеллектуальных возможностях, а с другой стороны, заведомое отчуждение человека от носителя искусственного интеллекта (Тьюринг, как убедительно описывает автор, в рамках своих тестов исходил из презумпции автономного действия машины). Такая презумпция, по большому счёту, и породила в массовом сознании представление о возможном «восстании

машин», о том, что искусственный интеллект может стать довлеющей и неуправляемой силой по отношению к конкретному человеку. Для обозначения якобы существующей непреодолимой границы между человеком и носителями искусственного интеллекта, автор диссертации вводит весьма меткое определение «стена Тьюринга».

Однако, как показывает тот же самый автор диссертации, не только в рамках предложенных им теоретических моделей, но и на материале непосредственно инженерных разработок его коллектива, эта «стена» является сугубо интеллектуальным конструктом, удобным для анализа систем «человек + техника» на определенном этапе их существования, но, во-первых, не имеющем прототипа в реальной действительности, а во-вторых, устаревшим даже в качестве методологического конструкта. Ключевым выводом автора является представление о продуктивном сосуществовании и взаимодействии искусственного и естественного интеллекта, позволяющем максимально реализовать потенциал первого и максимально усилить исходные интенции второго.

Стоит подчеркнуть, что автор ни в коей мере не говорит про вторичное, заведомо «подчиненное» положение искусственного интеллекта в сравнении с естественным (что было бы наиболее очевидным разрешением вопроса об их соотношении). Напротив, он показывает действительные возможности, при которых, искусственный интеллект приобретает отдельные элементы способности к творчеству и даже к построению действий в соответствии с приоритетами (насколько оно тождественно целеполаганию и, тем более, самоопределению, пока невозможно определить). При этом, данные возможности автор показывает в режиме структурно-генетического анализа, вводя категорию «техно-умвельта» как среза реального отражения мира, воспринимаемого устройством, которое управляется искусственным интеллектом. С точки зрения автора, программная разработка, способная к работе с двумя и более «техно-умвельтами», уже приближается по своим возможностям к естественному интеллекту. Ценно, что автор дает классификацию «техно-умвельтов» по очевидным критериям. Она, безусловно, нуждается в уточнении, но уже сейчас может служить методологической основой для конкретных программных разработок. Кроме того, модель «техно-умвельта», при значительной корректировке, может быть перенесена и в исследования, связанные с анализом «естественного» интеллекта.

К недостаткам представленной работы можно отнести сравнительно незначительную разработку таких категорий, которые, согласно базовым представлениям современной философской антропологии, принципиально отличают человека от машины при всем возможном равенстве интеллектуальных потенциалов: воли, наличие осознанных и бессознательных побуждений (интенций) структурировать мир в соответствии с собственными интересами, сама способность иметь имманентные интересы, и т.д. . Отчасти этот дефицит снимается базовым выводом автора о некорректности жесткого разделения искусственного и естественного интеллектов, об их заведомом переплетении и взаимодействии в процессе функционирования.

Диссертационное исследование выполнено методологически корректно, с высокой степенью оригинальности, опирается на большой объем проанализированной литературы, а также на концептуализированный опыт практической деятельности самого автора, содержит практически значимые выводы. Автор заслуживает присуждения ученой степени кандидата философских наук.

Александр Анатольевич Попов, доктор философских наук, генеральный директор Автономной некоммерческой организации дополнительного профессионального образования «Открытое образование», заведующий научно-исследовательским сектором «Открытое образование» Федерального института развития образования Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации

28 апреля 2021 год



Handwritten signature in blue ink.