

ОТЗЫВ

**официального оппонента доктора философских наук, профессора
Асеевой Ирины Александровны на диссертационную работу**

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**

Актуальность темы диссертации Альберта Рувимовича Ефимова определяется растущей значимостью интеллектуальной робототехники и искусственного интеллекта для решения прикладных научных задач и разработки новых технологий для развития современной цивилизации. Однако, все еще проблемным остается понимание конечной точки технологической дорожной карты – как мы узнаем, что действительно создали робота, который по своим способностям не только не уступает человеку, но и не может принести вреда отдельному человеку и человечеству в целом. А. Тьюринг на заре компьютерной эры предложил использовать «игру в имитацию» для оценки достижения машиной интеллекта человека. Однако, через 70 лет после его эпохальной статьи, эта точка зрения все еще оспаривается и широко обсуждается в мировой научной литературе. Это создает двойственность в отношении целеполагания по созданию автоматизированной системы с большой степенью функциональной автономии и искусственного интеллекта. Для преодоления этих теоретических и методических трудностей диссертант поставил цель своей работы – через комплексный анализ основных технологических трендов, современного состояния и научных перспектив интеллектуальной робототехники дать глубокий, философский анализ проблемы и предложить

возможный способ ее решения в оригинальной концепции посттьюринговой методологии.

Представленная на рецензию работа А.Р. Ефимова является, по моему мнению, глубоким и качественным исследованием, о чем свидетельствует обоснованная актуальность, подробная проработка проблемы, четко сформулированные объект, предмет, цель, задачи исследования, полученные теоретические результаты и успешная попытка проверить их в практическом использовании.

Диссертацию отличает грамотное аргументирование своей точки зрения и стройная логика рассуждений. Работа состоит из 9 параграфов, размещённых в трех главах, введения, заключения и списка литературы.

В первой главе «Методологический анализ современных исследований в интеллектуальной робототехнике» автор проводит глубокий и всесторонний обзор современного состояния вопроса в области научных исследований по направлению интеллектуальной робототехники, формирует систему понятий, в рамках которой строится все дальнейшее исследование: искусственный интеллект, Общий искусственный интеллект, интеллектуальная робототехника, мышление, сознание. Сильной стороной обзора является его подкрепленность примерами практического применения интеллектуальных роботов. Это дает возможность автору выработать рабочее определение интеллектуального робота, которое затем используется на протяжении всего исследования. Диссертант предлагает называть интеллектуальным роботом машину, обладающую 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 (стр.30). Обзор также является фундаментом для проведенной автором методологической дискуссии о значении для робототехники создания Общего искусственного интеллекта. Завершается первая глава постановкой эпистемологической проблемы субъектности роботов.

Вторая глава посвящена в основном анализу и осмыслению наследия А. Тьюринга и влиянию, которое оказал на развитие робототехники и искусственного интеллекта его знаменитый мысленный эксперимент «игра в имитацию». Автор справедливо отмечает, что А. Тьюринг сосредоточился на программном воплощении искусственного интеллекта исключительно из-за неразвитости аппаратного обеспечения. Основываясь на ряде оригинальных идей А.Тьюринга и проводя их развитие, автор вводит новую и интересную классификацию различных тестов, определяющих прогресс искусственного интеллекта.

Проблема идентификации достижения Общего искусственного интеллекта является сложной и актуальной. Автору удалось дать новый взгляд и ввести в научный оборот новые концепции, которые помогут в развитии новых когнитивных архитектур для интеллектуальной робототехники и Общего искусственного интеллекта.

Наиболее ценным в диссертации мне представляется развитие идей тьюринговой методологии и приложение ее к реальному физическому миру через призму четырех техно-умвельтов, которые вводит автор в своем исследовании. Концепция частных тестов Тьюринга развивалась в трудах А.Ю. Алексева, однако, автору удалось предложить новаторский подход к ее классификации, выделяя четыре измерения – виртуальное-вербальное (именно на нем был сосредоточен А. Тьюринг), виртуальное-невербальное, физическое-невербальное и физическое-вербальное. Такой подход к интеллектуальной робототехнике находит свое отражение во взглядах известных западных специалистов: инженера Р. Брукса и нейрофизиолога А. Дамасио.

Важнейшей мыслью диссертации является трансгрессия от тьюринговой методологии оценки достижений роботов или компьютеров с искусственным интеллектом к более приближенной к реальности пост-тьюринговой методологии. Здесь прослеживается явная связь с идеями постнеклассики. Представляется, что это оправданно подчеркивает фундаментальный переход от замещения роботом человека к взаимодействию робота и человека для реализации концепции Общего искусственного интеллекта.

Ценным в диссертации мне представляется развитие идей тьюринговой методологии и приложение ее к реальному физическому миру через призму четырех техно-умвельтов, которые выстраивает автор в своем исследовании. Автор опирается на концепцию умвельта, заложенную в работах Я. Иксюля, Е. Н. Князевой, В. Г. Буданова, и осмысление частных тестов Тьюринга в трудах А.Ю. Алексева. В продолжение этой темы автор диссертации для описания «жизненного мира» робота развивает понятие «техно-умвельт», зависящий от способности машины «видеть», адекватно отображать окружающий мир и действовать в меру своих возможностей (стр. 95). Диссертанту удалось предложить новый подход, выделить четыре измерения техно-умвельта – виртуальное-вербальное (именно на нем был сосредоточен А. Тьюринг), виртуальное-невербальное, физическое-невербальное и физическое-вербальное. Автор отмечает ограниченность любых попыток создать Общий искусственный интеллект без опоры на все четыре техно-умвельта. Этот подход является полноценным развитием идей, которые ранее выдвигались Д.И. Дубровским при его исследовании динамической структуры сознания человека.

Кроме указанных в диссертации новых идей мне представляются весьма интересными, но дискуссионными, гипотезы автора, который он развил в третьей главе:

1) неразличимость роботов и людей не является новой идеей сама по себе. Однако, автору удалось показать, что путь к неразличимости может

быть не только на основе создания интеллектуальных роботов, имеющих физическое воплощение, но и в создании виртуальных роботов, обладающих сходными с человеком характеристиками, но реализованными в виртуальном, компьютерном мире. Автор приводит в пример собственную разработку «Э.ЛЕНУ» - цифровую телеведущую, как способ реализации неразличимости;

2) влияние Общего искусственного интеллекта и интеллектуальной робототехники на образование. Технологии, подобные «Э.ЛЕНА» могут широко использоваться в образовательном процессе, дополняя учителя, преподавателя. Однако, такое использование должно быть тщательно продумано;

3) не менее значительным может быть влияние сочетания Общего искусственного интеллекта и технологии виртуальных deep-fakes на индустрию развлечения. Автор предполагает, что уже сейчас или в недалеком будущем можно будет виртуально оживить давно ушедших актеров для того, чтобы они могли сыграть роли, которые никогда в их жизни не присутствовали. Остается, правда, открытой этическая спорность такой технической возможности.

Таким образом, теоретическая значимость работы заключается в том, что диссертанту удалось убедительно обосновать посттьюринговую методологию для развития интеллектуальной робототехники и создать методологическую основу для классификации частных тестов Тьюринга. В целом содержание диссертационной работы отражает поставленные автором задачи, соответствует теоретико-методологической направленности и последовательно раскрывает вопросы в содержании глав и параграфов.

Наряду с достижениями автора необходимо отметить и ряд недоработок, которые надо учесть в дальнейших исследованиях.

1. В диссертации не в полной мере раскрыт вопрос о различии между понятиями «сильный» и «общий» искусственный интеллект. Известно, что Джон Сёрл ввел термин «сильный искусственный интеллект»

в своей статье 1980 г. Термин «общий искусственный интеллект» получил достаточно широкое распространение лишь в последнее десятилетие. Необходимо более серьезное исследование обоснованности использования этого термина в контексте философии искусственного интеллекта и корректное его использование с учетом смысловых нюансов.

2. Вопрос о связи посттюринговой методологии с постнеклассикой заслуживает отдельного глубокого изучения и может стать самостоятельным исследованием. Наши философы науки (В.С. Степин, В.И. Аршинов, В.Г. Буданов, И.В. Черникова) имеют богатый опыт осмысления современной физики через постнеклассический подход. Перенесение этого опыта на актуальную тему изучения искусственного интеллекта могло бы быть довольно плодотворным.

3. В первой главе затрагивается сложнейшая эпистемологическая и этическая проблема субъектности машин с искусственным интеллектом. Робот может быть оснащен сложным программным обеспечением, которое будет принимать информацию и выполнять вычисления, помогающие выбрать адекватную реакцию на ситуацию, в которой он находится, но для наличия субъектности этого не достаточно. Будет ли у робота сознательный субъективный опыт, эмпатия, осознанные ценности и добровольный выбор? Или согласиться с Джоанной Брайсон, что лучше не делать слишком продвинутых роботов, чтобы избежать тревожной двусмысленности в отношении их морального статуса? Проблема «метафизических» характеристик робота в данной диссертации фактически не разработана. Между тем, с моей точки зрения, философско-этическая рефлексия создания интеллектуальных роботов должна быть вписана в технологический процесс уже на этапе поиска методологии.

4. Введенный автором термин и понятие «техно-умвельта» требует более тщательной разработки в контексте своих связей с другими возможными применениями идеи «умвельта». Автору следует более широко

посмотреть на генезис умвельта и провести параллели с социумом, кибернетическими и управленческими практиками.

5. Необходимо отметить, что диссертация и исследования автора основываются на серьезном практическом опыте работы, который подтвержден не только научными статьями, но и патентами на изобретения и внедрением их в повседневную жизнь. Прототип «Э.ЛЕНЫ», авторского теста для проверки качества искусственного интеллекта, разработанный коллективом под руководством А.Р. Ефимова уже используется в экосистеме Сбербанка для автоматической генерации видеоконтента для вещания. Однако диссертант скромно упоминает его лишь в ряде других, конкретного анализа этого теста не приводится, не аргументируются его преимущества или потенциальные возможности. Представляется, что такой тест на искусственный интеллект, обладающий визуальной репрезентацией, был бы хорошим дополнением к тестам, классифицированных автором.

Вместе с тем, высказанные замечания обозначены в порядке научной дискуссии и не снижают общей высокой положительной оценки представленной работы.

Область исследования диссертации Ефимова Альберта Рувимовича на тему: «Философско-методологические основы посттюринговой интеллектуальной робототехники» соответствует номенклатуре специальностей научных работников 09.00.08 – философия науки и техники в п.3 - Значение социальных факторов для прогресса науки и техники; п.9 - Анализ основных тенденций развития современной науки и техники; п.20 - Гуманизация и гуманитаризация современной науки и техники; п.27 - Философско-методологические аспекты информатики и информатизации общества.

Автореферат объективно и в полной мере отражает содержание диссертации. Структурное построение автореферата диссертации отвечает предъявляемым требованиям, в нём присутствуют все необходимые установочно-содержательные элементы.

Основные идеи исследования прошли достаточную научную апробацию в рамках различных научно-практических мероприятий и нашли отражение в 11 статьях, из них 3 - в журналах, рекомендованных ВАК.

Резюмируя сказанное, можно сделать вывод о том, что диссертация Альберта Рувимовича Ефимова является самостоятельной завершенной квалификационной научной работой. Ее содержание отвечает требованиям ВАК РФ, предъявляемым к диссертационным работам на соискание учёной степени кандидата наук и соответствует необходимым требованиям п.п. 9 и 10 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 года № 842, а ее автор заслуживает присуждения ему искомой степени кандидата философских наук по специальности 09.00.08 – философия науки и техники.

27 апреля 2021 года

Официальный оппонент
Доктор философских наук
(09.00.08 – философия науки и техники), профессор,
ведущий научный сотрудник Федерального государственного бюджетного учреждения науки Института научной информации по общественным наукам Российской академии наук



Асеева Ирина Александровна

Институт научной информации по общественным наукам РАН
117418, Москва, Нахимовский проспект, д. 51/21
<http://inion.ru>
тел. +7(499)124-37-83
office-inion@mail.ru

