



КОГНИТИВНОЕ ПРОСТРАНСТВО



Философия сознания



Снятся ли чат-ботам андройды? Перспективы технологического развития искусственного интеллекта и робототехники

А.Р. Ефимов

Лаборатория робототехники ПАО «Сбербанк», Москва, Россия

DOI: 10.30727/0235-1188-2019-62-7-73-95

Оригинальная исследовательская статья

Аннотация

Статья посвящена обобщению основных трендов развития систем искусственного интеллекта и робототехники (ИИиР). Основной вопрос, который рассматривается в этом контексте: будут ли искусственные системы становиться все более антропоморфными как в интеллектуальном, так и в физическом отношении? В статье автор не только проводит анализ современного состояния и перспектив технологического развития искусственного интеллекта и робототехники, но и определяет основные аспекты влияния этих технологий на общество и экономику, указывая на геополитический стратегический характер данного влияния. Автор рассматривает различные подходы к определению искусственного интеллекта и робототехники, выделяя предметно-ориентированный и функциональный. Также производится сопоставление способностей ИИиР и человеческих способностей в таких областях, как категоризация, распознавание образов, планирование и принятие решений и др. На основе этого сопоставления сделаны выводы о том, когда ИИиР уступают человеку, а в каких случаях превосходят его. Проанализированные автором современные достижения в области робототехники и искусственного интеллекта создают необходимый базис для дальнейшего рассуждения о применимости инженерного целеполагания в виде теста Тьюринга. Показано, что развитие ИИиР связано с определенными противоречиями, затрудняющими применение методологии Тьюринга в привычном формате. Рассмотренные автором базовые противоречия в развитии технологий ИИиР позволяют сделать основной вывод о переходе к посттьюринговой методологии оценки инженерных реализаций

искусственного интеллекта и робототехники, в которых, с одной стороны, снимается «стена Тьюринга», а с другой стороны, искусственный интеллект получает свое физическое воплощение.

Ключевые слова: робототехника, искусственный интеллект, Тьюринг, тест Тьюринга, философия искусственного интеллекта.

Ефимов Альберт Рувимович – руководитель Лаборатории робототехники ПАО «Сбербанк», аспирант философского факультета МГУ имени М.В. Ломоносова.

makkawity@gmail.com

<https://orcid.org/0000-0001-6857-8659>

Для цитирования: Ефимов А.Р. (2019) Снятся ли чат-ботам андроиды? Перспективы технологического развития искусственного интеллекта и робототехники // Философские науки. 2019. Т. 62. № 7. С. 73–95. DOI: 10.30727/0235-1188-2019-62-7-73-95

Do Chatbots Dream of Androids? Prospects for the Technological Development of Artificial Intelligence and Robotics

A.R. Efimov

Robotics Laboratory of PJSC Sberbank, Moscow, Russia

DOI: 10.30727/0235-1188-2019-62-7-73-95

Original research paper

Abstract

The article discusses the main trends in the development of artificial intelligence systems and robotics (AI&R). The main question that is considered in this context is whether artificial systems are going to become more and more anthropomorphic, both intellectually and physically. In the article, the author analyzes the current state and prospects of technological development of artificial intelligence and robotics and also determines the main aspects of the impact of these technologies on society and economy, indicating the geopolitical strategic nature of this influence. The author considers various approaches to the definition of artificial intelligence and robotics, focusing on the subject-oriented and functional ones. In the article, AI&R abilities and human abilities are compared in such areas as categorization, pattern recognition, planning and decision making, etc. Based on this comparison, it is concluded in which areas AI&R's performance is inferior to a human and in which cases it is superior to one. The modern achievements in the field of robotics and artificial intelligence create the

necessary basis for further discussion of the applicability of goal setting in engineering, in the form of a Turing test. It is shown that development of AI&R is associated with certain contradictions that impede the application of Turing's methodology in its usual format. The basic contradictions in the development of AI&R technologies imply that there is to be a transition to a post-Turing methodology for assessing engineering implementations of AI&R. In such implementations, on the one hand, the "Turing wall" is removed, and on the other hand, artificial intelligence gets its physical implementation.

Keywords: robotics, artificial intelligence, Turing, Turing test, philosophy of artificial intelligence.

Albert R. Efimov – Head of PJSC Sberbank's Robotics Laboratory, postgraduate student at the Faculty of Philosophy, Lomonosov Moscow State University.

makkawity@gmail.com,

<https://orcid.org/0000-0001-6857-8659>

For citation: Efimov A.R. (2019) Do Chatbots Dream of Androids? Prospects for the Technological Development of Artificial Intelligence and Robotics. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 62, no. 7, pp. 73–95. DOI: 10.30727/0235-1188-2019-62-7-73-95

Введение

Искусственный интеллект и робототехника являются одними из самых молодых областей науки и технологий. Несмотря на давние традиции технологического антропоморфизма и органопроекции, восходящие еще к древнегреческим философам Протагору и Сократу [Дрейфус 1978], подлинная история машин, способных хотя бы отдаленно рассуждать или действовать подобно человеку, насчитывает всего лишь несколько десятилетий. Ведь семьдесят лет назад «вычислитель» (computer) еще был должностью в штатном расписании многих организаций.

В течение короткой истории компьютерных технологий, кибернетики, информатики и робототехники не было недостатка в рекламных лозунгах, воспевающих их преимущества и эффективность по сравнению с предыдущими инструментами ведения хозяйства. В стремлении привлечь внимание как можно большего числа потребителей резко меняются и их названия – от кибернетики до искусственного интеллекта. Однако, следуя меткому замечанию нобелевского лауреата Роберта Солоу, которое он сделал более 30 лет назад, «влияние компьютерной эпохи заметно

езде, за исключением статистики производительности труда» [Solow 1987, 36].

За прошедшие три десятилетия технологии сделали значительный шаг вперед, но именно сейчас большинство экономистов согласны с тем, что искусственный интеллект и робототехника (в дальнейшем – ИИиР) будут оказывать существенное влияние на нашу цивилизацию. Впрочем, не только экономика определяет повышенное внимание суверенных государств и правительств к возможностям, которыми обладают ИИиР. Эти технологии становятся полноценным инструментом геополитики, с помощью которого победители получают все, а проигравшие будут вынуждены жить по правилам, установленным победителями [Shaw 2017].

Научная фантастика предоставляет огромный выбор тем как для кинорежиссеров, так и для философов. Даже название данной статьи является перефразированием заголовка известного романа Ф. Дика «Сняты ли андроидом электрические овцы?» (1), целиком посвященного сложной проблеме возникновения сознания на небиологическом субстрате. Однако данная статья посвящена некоторым аспектам взаимодействия искусственного интеллекта (2) (или в общем случае программного обеспечения, «психики») и роботов (или в общем случае аппаратного обеспечения, носителя для искусственного интеллекта, «тела», «моторики») (3).

Определения искусственного интеллекта и робототехники

Вследствие того, что актуальность тематики ИИиР значительно возросла, дать точное определение ИИ – это крайне сложная проблема. Представители академической науки и бизнеса не имеют единой точки зрения на то, что такое ИИиР. Однако отсутствие единого определения не мешает росту инвестиционной и деловой активности, что, в свою очередь, делает еще более актуальной необходимость предложить бизнесу более адекватные способы обозначения предметной области ИИиР с помощью собственного понятийного аппарата.

Наиболее точной представляется модель, предложенная В.К. Финном, согласно которой ИИ следует интерпретировать через определение естественного интеллекта, имеющего следующие характеристики, реализуемые на небиологическом субстрате (математическом обеспечении, алгоритмах и аппаратной реализации):

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

2. способность к целеполаганию и планированию поведения – порождение последовательностей “цель → план → действие”;

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

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

5. способность к аргументированному принятию решений, когда используются упорядоченные знания (представление знаний) и результаты рассуждений, соответствующие поставленной цели;

6. способность к рефлексии – оценке знаний и действий;

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

8. способность и потребность находить объяснение (не обязательно дедуктивное!) как ответ на вопрос “почему?”;

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

10. способность к обучению и использованию памяти;

11. способность к рационализации идей: стремление уточнить их как понятия;

12. способность к созданию целостной картины относительно предмета мышления, объединяющей знания, релевантные поставленной цели (т.е. формирование по крайней мере приближенной теории предметной области);

13. способность к адаптации в условиях изменения жизненных ситуаций и знаний, что означает коррекцию “теорий” и поведения» [Финн 2018, 37].

По мнению Финна, способности, обозначенные в пп. 1, 3, 4, 5, 6, 8, 9, могут быть реализованы машинами без участия человека при наличии достаточных программных и аппаратных средств, но способности, представленные в пп. 2, 7, 12 и 13 не могут быть реализованы без частичного участия человека, взаимодействующего с машиной.

Само понятие «робототехника» сложноопределяемо, хотя в индустриальной робототехнике уже существует глобальный стандарт, определяющий понятие робота: «Робот – это рабочий механизм, программируемый по нескольким степеням свободы с некоторой степенью автономности и способный передвигаться в пределах определенной среды, выполняя поставленные задачи» [ISO 8373:2012].

Недостаток данного определения становится очевиден, если задуматься о том, что под него подпадают совершенно разные объекты, такие как беспилотные летательные аппараты, самоуправляемые автомобили и промышленные манипуляторы со многими степенями свободы. При этом данное определение не имеет никакой корреляции с рядом передовых исследований, которые говорят о том, что роботы будущего – это, прежде всего, инструменты, способные к групповому взаимодействию друг с другом, самоорганизации и общению с человеком на естественном языке.

Подобное отставание определений ИИиР от технического прогресса дает импульс для простых инженерных дефиниций исследуемых областей. В ИИиР это происходит по примеру А. Тьюринга, который подошел к определению ИИ (посредством вопроса «может ли машина мыслить?») через антиэссенциалистский отказ глубоко исследовать «сущность» ИИ и постановку инженерной задачи. В частности, отказываясь от малоэффективных онтологических споров о том, что такое ИИ, эксперты Сбербанка рассматривают ИИ как алгоритмы, программное и аппаратное обеспечение, решающие следующие прикладные задачи на уровне человека или выше:

- компьютерное зрение;
- рекомендательные системы и интеллектуальные системы поддержки принятия решений;
- обработка естественного языка;
- распознавание и синтез речи;
- перспективные методы и технологии ИИ.

Автором было предложено *прикладное* функционалистское определение робототехники, которое основано на адаптации «утиного теста» (4). Согласно этому определению, роботом можно назвать любое устройство, которое способно одновременно:

1. *Sense*: робот обладает способностью воспринимать окружающий мир с помощью сенсоров. Такими сенсорами могут быть микрофоны

(сонары), камеры (всех областей электромагнитного спектра), различные электромеханические сенсоры (акселерометр) и прочее.

2. *Think*: робот обладает способностью интерпретировать (понимать) сигналы, которые он получает от сенсоров, наблюдающих физический мир, строить и адаптировать модели поведения и принимать решения в зависимости от выбранных моделей. Эта способность может быть реализована разными способами: бортовым вычислителем робота, «интеллектуальным» облаком или человеком, который управляет роботом с помощью телеуправления или интерфейса с реализованной обратной связью.

3. *Act*: робот обладает способностью воздействовать на физический мир любым результативным способом [Ефимов и др. 2019].

Подходы специалистов Сбербанка, включая предложенный автором, являются, по сути, узким применением философского приема М. Хайдеггера, который перешел от платоновской рациональной абстракции к «эмоционально-практическому» отношению к миру вообще и к использованию техники и технологий ИИиР в частности [Dreyfus 1991, 8].

По выражению Д. Маккарти, словосочетание «искусственный интеллект» является «словом-чемоданом», т.е. словом, имеющим невероятно большое количество значений. Наверное, сейчас количество возможных значений уже равно числу всевозможных подходов к исследованиям или коммерциализации технологий, что делает бессмысленными статические, формально-логические определения «искусственного интеллекта» и заставляет исследователей и инженеров искать более релевантные способы описания собственных областей деятельности. Одним из таких подходов как раз может стать принятый в Сбербанке подход для описания ИИ как набора областей применения или робототехники как ряда критических функций. Исследователи и инженеры, действуя в собственных интересах и движимые осознанием возможности реализации себя через предлагаемые ими технические решения или научные открытия, могут полагаться на такие практические определения и иметь устойчивую эмоциональную связь с ясно описанной предметной или функциональной областью.

Текущие и перспективные области применения искусственного интеллекта и робототехники

С начала промышленной революции те отрасли человеческой деятельности, в которых автоматизация была применима,

находились под пристальным вниманием предпринимателей и инженеров. Любую механическую работу, которую можно было автоматизировать на текущем уровне развития техники, обычно старались автоматизировать при условии того, что это приносит значительную экономию ресурсов. При этом отрасли, где деятельность человека была основана на правилах или знаниях, были во многом исключены из-под «обесчеловечивающего» влияния технологий. «Если обратиться к предложенному Расмуссеном различению видов деятельности на основанную на (а) навыках (skill-based), (b) правилах (rule-based), (c) знании (knowledge-based), то можно увидеть, что автоматизация касалась в основном первого и второго видов деятельности [Rasmussen 1979].

ИИ обладает уникальным свойством превращения областей деятельности, которые ранее не подвергались влиянию автоматизации и были уделом «творческих» людей (т.е. (c) по Rasmussen), в такие же технические отрасли, как и те, где автоматизация уже активно применяется столетиями. И происходит это благодаря возможности накапливать данные в цифровом виде. Юриспруденция, образование, финансы, медицина и экономика постепенно превращаются в наборы алгоритмов и данных, а деятельность людей в этих отраслях все больше становится близкой к ремеслу (а не к умению), к тому, что древние греки называли *βαναυσία*, которая превращает мастерское умение типа (c) в банальную работу типа (a).

Многие экономисты видят в ИИиР возможности для роста общей факторной производительности труда, которая стабильно снижается с 80-х годов прошлого века, несмотря на переход нашей цивилизации к информационной эпохе (см., например: [Бриньолфсон и др. 2019]). Текущие экономические оценки дают рост производительности труда на 40% в целом по экономике за счет использования ИИиР. В случае полномасштабного проникновения технологий, основанных на ИИиР, в экономику, экономический рост, обусловленный их внедрением, будет весьма значительным, сравнимым по эффекту с появлением паровой машины в первой индустриальной революции XVIII в. В экономически развитых странах дополнительный рост, связанный с ИИиР, может составлять от 1,8% до 4,6% ВВП ежегодно [Purdy et al. 2016].

На самом деле, ИИиР является типичной сквозной технологией, применяемой в любой сфере народного хозяйства и государственного управления. В связи с этим стоит указать лишь наиболее

значимые возможности для применения ИИиР. Для робототехники к таким сферам относятся: обработка изображений, обработка данных сенсоров, предиктивный анализ и моделирование, голосовые помощники, чат-боты, логистика, розничная торговля, сельское хозяйство, уход за пожилыми. Что касается искусственного интеллекта, помимо всего перечисленного, эти технологии применяются в образовании, синтезировании изображений и виртуальных миров, анализе данных сенсоров, анализе медицинских данных.

Некоторые значимые достижения искусственного интеллекта и робототехники за последние 20 лет

Основоположник квантовых вычислений Дэвид Дойч справедливо замечает, что любые инженерные изыскания в области общего ИИ, не поддержанные серьезной эпистемологией и методическими основаниями, будут бесплодными [Deutsch 2012]. Однако в том случае, если философская наука и ее представители сами не погружены в новейшие технологические достижения и не осознают текущие и будущие возможности, а также ограничения, то возникает недооценка перспектив ИИиР или же, наоборот, появляются завышенные ожидания. В данном разделе делается попытка преодолеть разрыв между инженерными достижениями последних лет и их философским осмыслением.

Для начала ответим на вопрос о том, где и в чем искусственный интеллект и роботы лучше человека. В самом начале эры информационных технологий Алан Тьюринг положил начало сравнению человека и компьютера, заложив тем самым основы машинного функционализма [Putnam 1975]. Продолжая эту традицию, Дж. Фодор [Fodor 1990] заложил основы компьютеризма, который был в дальнейшем развит в тестовый функционализм [Алексеев 2013]. Во многом вся дальнейшая история ИИиР стала соревнованием, в котором одни исследователи убедительно доказывали, что машина никогда не сумеет мыслить подобно человеку [Дрейфус 1978], а другие – убедительно показывали, что это не только возможно, но и уже достигнуто [Krol 1999]. Однако общее сопоставление достижений ИИиР в сравнении с возможностями человека пока еще не в пользу машин.

Рассмотрим некоторые из областей, в которых можно сравнить ИИиР и человека. Здесь мы пойдем дальше традиционного сопоставления естественных и искусственных логико-лингвистических

способностей. Нас интересуют также психомоторные функции и пределы их реализации в искусственных системах. Это связано с тем, что интеллектуальная деятельность обладает психомоторным базисом. Так, общеизвестно, что развитие моторных функций у детей является одним из ключевых факторов для дальнейшего развития интеллекта в целом. Это подтверждается и общей эволюцией, которая демонстрирует теснейшую связь моторных и психических функций. Именно поэтому возникновение и развитие психики произошло именно у тех сложных организмов, которые активно передвигаются во внешней среде [Dubrovsky 2018].

Сопоставляя способности ИИиР и человека, мы пришли к выводу о том, что человек пока еще превосходит искусственные системы по большинству пунктов. Так, например, человек превосходит ИИиР в области контекстуализации взаимоотношений между вещами, предсказания для понимания причинно-следственной связи, планирования и принятия решений, мелкой моторики, динамики и движения, поднятия тяжестей, энергоэффективности. Уступает же человек в двух аспектах: восприятие и категоризация окружающего мира и точность (повторяемость) движений. Согласно данным, представленным в исследовании «AI Index» [Shoham et al. 2018], точность распознавания категорий образов с помощью систем на основе нейронных сетей глубокого обучения статистически превзошла уровень человека в 2015 г. В настоящее время наблюдается дальнейший устойчивый прогресс в качестве распознавания образов. Сравнительная точность (повторяемость) движения человека и роботов измеряется в биомеханике и эргономике, но в целом можно сказать, что ошибка в движении современных роботов при последующих движениях к той же точке из того же направления и при тех же условиях не превышает 0,1 мм при условии тысяч повторений. Подобной точности человек достигнуть не может хотя бы потому, что ему нужны перерывы для отдыха, сна, еды и т.п.

Инженерные воплощения теста Тьюринга (2014–2019 гг.)

Переход от абстрактных мечтаний к инженерным решениям не раз помогал человечеству в преодолении знаковых барьеров. К примеру, на заре развития авиации именно рекордные перелеты через Ла-Манш, Атлантический и Северный Ледовитый океан сделали транспорт (а косвенно и всю нашу современную цивилизацию) именно такой, какой мы ее сейчас знаем. Алан

Тьюринг, возможно, не проводил подобной параллели, предлагая «игру в имитацию» для определения мыслительных способностей машин, однако произошедшее транспонирование его идеи на технологическое развитие было неизбежным. В мире проводится ряд технологических конкурсов, направленных на проверку реализации ТТ и основанных на игре в имитацию. Конкурсы можно разделить на два типа: классические (в них реализуется ТТ типа 2 [Алексеев 2013; Warwick, Shah 2016]) и неклассические (в них реализуются различные модификации ТТ).

Образцом классического подхода к ТТ служит эксперимент, проведенный К. Варвиком и Х. Ша в Королевском обществе в Лондоне в 2014 г. и описанный в различных работах (к примеру, [Warwick, Shah 2016]). Также существует критика этого подхода к проведению теста, еще раньше представленная в ряде работ, например в коллективной монографии «The Turing Test: The Elusive Standard of Artificial Intelligence» [Moor 2003]. Несмотря на критику подхода, тест, проведенный в 2014 г., в точности воспроизводил условия, представленные А. Тьюрингом в его работе «Вычислительные машины и разум» [Turing 1950]. Упрощенно эти условия следующие (терминология по [Алексеев 2013]): судья (С.) должен неправильно идентифицировать человека (Ч.) или компьютер (К.) в более чем 30% случаев после диалога, который длится пять минут. В тесте участвовало больше 30 исследователей, включая автора в качестве С. В результате чат-бот К. Eugene Goostman, имитирующий подростка из Одессы, сумел обмануть 33% С. в сравнении с Ч. Несмотря на то, что это достижение подвергалось во многом справедливой критике, его следует признать значимым, поскольку оно косвенно свидетельствует о неактуальности классического подхода и буквальной бихевиористской интерпретации теста Тьюринга. Ведь прохождение теста не приводит к новым прорывам в области развития искусственного интеллекта, тогда как, например, в случае трансатлантических перелетов начала XX в. или космических полетов середины того же века решение ключевых для отрасли проблем привело к прорыву в новых областях науки и техники.

Неклассические интерпретации теста Тьюринга, такие как тесты, проведенные в 2015 г. в Сколково и в 2019 г. на конференции «OpenTalks.AI», являются больше рекламными акциями, чем попыткой определить возможности машины для диалога с человеком. В частности, авторы теста 2019 г. подчеркивают, что

в ходе диалогов Ч. очень часто сознательно имитировали стиль диалога чат-ботов, чтобы ввести в заблуждение С. По сути, это не повышение интеллекта у машины, но сознательное снижение интеллекта у Ч. в глазах С. [Наносемантика 2019]. Кроме того, вместо пяти минут общения организаторы ограничили возможность С. для общения с Ч. или К. лишь двенадцатью репликами, что явно недостаточно для распознавания К. или Ч. в диалоге. Именно этими двумя обстоятельствами и объясняется то, что порог обмана в этом тесте составил 58% (т.е. в 58% К. удалось выдать свою личность за Ч. в глазах С.). Иными словами, если Ч. при общении с ИИиР перестает общаться привычным ему образом и начинает говорить в командно-приказной модальности речи, то ИИиР парадоксальным образом начинает лучше понимать человека, а человек, сознавая, что он разговаривает с чат-ботом, испытывает гораздо меньше разочарования, чем при разговоре так, как если бы чат-бот (робот) был человеком и понимал его как человек. Интуитивно можно предположить, что мы находимся в самом начале зарождения особого упрощенного языка общения человека и машины, который носит очень комбинированный характер, сочетая лингвистический (слова и их смысл) и нелингвистический язык (звуки, световые сигналы). Это предположение само по себе может стать объектом очень интересного исследования, но должно быть более тщательно проработано.

Если вы боитесь восстания роботов, то просто закройте перед ними дверь (DARPA Robotics Challenge)

Бихевиористский подход к оценке достижений современной робототехники остается ее текущей интеллектуальной парадигмой, где присутствует стремление ответить на вопрос о том, может ли робот выполнять действия на уровне или лучше человека. Это оправдано, т.к. одной из главных причин внедрения робототехники является сохранение человеческих жизней при ликвидации стихийных бедствий, например ядерных катастроф. Авария и последующие трагические события на АЭС «Фукусима» в 2011 г. в Японии послужили предпосылкой для запуска технологического конкурса Robotics Challenge, проводимого американским агентством перспективных оборонных исследований DARPA [Krotkov et al. 2016]. Суть задачи состояла в том, что роботы должны выполнить действия, аналогичные тем, которые выполняют спасатели при ликвидации стихийных бедствий: доехать

самостоятельно на автомобиле, выйти из него, открыть дверь, зайти в комнату, найти инструмент, выполнить действия с ним, пройти через завал и взойти на лестницу. Для любого взрослого человека эти действия не составляют никакого труда и полоса препятствий, сконструированная организаторами, может быть пройдена человеком без какой-либо подготовки за считанные минуты. Финал конкурса был проведен в 2015 г. в США (Калифорния) и показал, что роботам нужно значительно более 40 минут (лучший результат – 44 минуты), чтобы успешно выполнить все девять поставленных задач. Основными выводами о статусе робототехники по результатам проведенного инженерного конкурса являются следующие:

Во-первых, мелкая моторика и локомоция роботов не превосходят уровня годовалого ребенка. Сложнейшими задачами для всех роботов были открывание двери и восхождение по лестнице. Эти задачи не вызывают проблем у всех здоровых людей, но для роботов они представляются невероятно сложными. Именно поэтому один журналист дал совет: если вас преследует робот, «существует эффективная тактика – закрывайте за собой дверь» [Hernandez 2017].

Во-вторых, роботы, участвовавшие в конкурсе, имели очень ограниченную автономность и большинство времени находились в режиме телеуправления, т.е., по сути, не отличались от радиоуправляемых игрушек или более сложных машин. Самостоятельно, без помощи людей, роботы были неспособны выполнить ни одного задания. Это ставит вопрос о необходимости повышения степени автономности и интеллектуальности роботов при проведении их испытаний в реальных условиях, а не во время студийных съемок (где обычно все хорошо).

Базовые противоречия развития технологий ИИиР

В своем развитии любая технология движется вдоль направлений базовых противоречий или т.н. «несущих противоречий» [Бескаравайный 2018]. В развитии ИИиР также можно увидеть сразу несколько направлений технологических противоречий. Автор не предлагает исчерпывающего списка, однако каждое из направлений можно представить в качестве вектора развития и возможной борьбы между конкурирующими технологиями. Иногда такие базовые противоречия называют противоборством «брони и снаряда» (5) [Бескаравайный 2018]. Продолжая эту метафору,

назовем ряд технологических противоречий ИИиР, представив их в виде «брони и снаряда».

Психика или моторика. Со стороны «брони» выступает *психика* – ИИиР должны постоянно улучшать сенсоры и программное обеспечение для того, чтобы лучше понимать окружающий мир. Это позволяет ИИиР строить адекватные модели и принимать более быстрые и эффективные решения. «*Снарядом*» здесь выступает *моторика* – роботы должны научиться передвигаться в окружающем мире так же, как это делают люди и животные для того, чтобы познавать объекты мира и изучать их, постоянно улучшая свои навыки восприятия. Без тонкой моторики для манипуляций роботы не могут ощущать мир подобно людям. Противостояние «психики» и «моторики» роботов не является новостью в робототехнике, оно известно как парадокс Моравека [Moravec 1988]. Смысл данного парадокса состоит в том, что относительно легко достичь уровня взрослого человека в таких задачах, как тест на интеллект или игра в шашки, однако сложно или невозможно достичь навыков годовалого ребенка в задачах восприятия или мобильности.

Компьютерная мощность или логические правила. Со стороны «брони» может быть компьютерная мощность, т.к. необходимо постоянно увеличивать объем компьютерной мощности, доступной для обработки глубоких нейронных сетей. По мнению исследователей проекта «OpenAI», рост процессорной мощности, используемой для тренировки нейронных сетей, превышает темпы роста процессорной мощности по закону Мура примерно в пять раз, а поколение систем 2019 г. превосходит по мощности поколение систем 2012 г. в 300 тысяч раз. «*Снарядом*» же здесь являются логические правила, которые можно использовать для создания программ, имитирующих интеллект, поскольку главной проблемой использования нейронных сетей и иных статистических методов является рациональное объяснение полученных результатов. Многие исследования указывают на то, что наиболее эффективным способом достижения ИИ общего уровня (т.н. AGI) будет создание комбинированных решений, сочетающих в себе логические подходы и нейронные сети глубокого обучения [Brooks 2019]. Необходимо отметить, что закон Мура в том виде, в каком он управлял всей компьютерной отраслью в течение почти 60 лет, подошел к своему концу. Создание технологий ИИиР безусловно опирается на микропроцессорные технологии,

однако дальнейшие прорывные исследования будут выполнены не благодаря избыточной компьютерной мощности, а за счет различных способов повышения эффективности вычислений вне чисто аппаратных решений.

Цифровое цунами vs цифровой фильтр. Со стороны «брони» в данном случае выступает *цифровое цунами*. Суть в том, что широко распространено заблуждение об экспоненциальном росте количества накапливаемой человечеством информации. Если оно и растет, то лишь линейно, с ростом самого человечества. Однако экспоненциально растет и способность человечества преобразовывать свои впечатления от окружающего мира в цифровой формат, тем самым создавая цифровое облако вокруг физического мира (который в целом не меняется сильно). В случае *цифрового фильтра «снарядом»* является упорядочение информации (фильтрация), которое может быть как функцией человека, так и миссией компании (Google). Проблему фильтрации необходимой информации можно также переформулировать как проблему обучения на малых данных, подобно тому, как обучаются маленькие дети, которым необходимо лишь один раз показать или рассказать, и они способны повторить пример без больших вычислительных и, следовательно, энергетических затрат.

Симбиоз человека и машины. Усложнение технологий и ускорение жизненных циклов (ЖЦ) наблюдается во всех областях развития технологий, включая ИИиР. «Броней» здесь является главный способ адаптации человека к изменениям – обучение. Работники практически любых специальностей вынуждены проходить через постоянное повышение квалификации. Обучение человека не меняется с каменного века и с момента появления речи. *Киборги* – люди, имеющие физическую интеграцию своих тел и машин, живущие буквально в симбиозе с машиной, являются единственным «снарядом», способом преодолеть пропасть, разделяющую человека и умные вещи, созданные им. Наиболее значимым технологическим направлением является развитие интерфейсов «мозг – компьютер», которые позволяют не только читать мысли человека, но и передавать новую информацию в мозг с огромной скоростью.

Тест Тьюринга и посттьюринговая эпоха

В своей работе, посвященной проблеме *другого* и тесту Тьюринга, А.Ю. Алексеев приводит выдуманную историю «Девушка

по переписке», которую хотелось бы упомянуть также здесь для того, чтобы с ее помощью проиллюстрировать общую идею. Суть истории в следующем. Молодой человек ищет спутницу жизни и, как сейчас принято, открывает приложение Tinder на своем смартфоне. Задает параметры запроса – возраст, рост, объемы, цвет волос, знак зодиака, привычки и пр. В ответ – ряд кандидатов. Начинается переписка. Особенно нравится девушка Элиза. Переписка длится целую неделю – в выборе нельзя ошибиться! Беседы ведутся на самые разные темы: семьи, детей, секса, быта, отдыха, кухни, зарплаты, экономики, политики, литературы и пр. Играют в шахматы с переменным успехом. Девушка поражает объемом знаний, легкостью в общении, изысканностью стиля. И глаза у нее выразительные, судя по присланной фотографии. Наконец, молодой человек не выдерживает и назначает встречу на площади у памятника. Приезжает с букетом роз. Ждет. Час, два, три. Никто не приходит. Вернувшись ни с чем, пишет гневное письмо. В ответ – «Прости, милый! Я – Электронная Лиза, компьютерная программа» [Алексеев 2008] (6).

Идею посттьюринговой эпохи очень легко проиллюстрировать на этом примере. Единственное изменение в том, что девушка по переписке *пришла* на свидание. Молодой человек отвел ее в кафе, и там они начали разговаривать. Разговор не складывался. Никаких следов образования, юмора или изысканности стиля в разговоре с девушкой не было. Глубоко разочарованный молодой человек вышел из кафе и открыл приложение, в котором была переписка с девушкой. Он открыл ее профиль, зашел в детальное описание и лишь тогда заметил маленькую галочку «пользователь включил возможность автоответов с помощью *цифрового аватара*». Только сейчас он осознал, что девушка, с которой он флиртовал всю неделю, даже никогда ему не отвечала и не читала его сообщения. Все это делал ее цифровой двойник, который был намного умнее и буквально знал, что хочет слышать молодой человек, т.к. система имела доступ к его профилю и могла адаптировать любые ответы под его запросы.

Тест Тьюринга можно трактовать как сфокусированный на ИИиР *утиный* тест: «если что-то говорит как человек, действует как человек, ведет себя как человек, то это человек». Но ответы на него могут быть принципиально разными. В частности, А.Ю. Алексеев и Д.И. Дубровский предлагают разные идеи о возможности прохождения теста Тьюринга. По мнению Алексеева,

если компьютер проходит ТТ, то этот ИИ в лучшем случае будет компьютерным зомби [Алексеев 2008]. По мнению Дубровского, прохождение ТТ немыслимо без переноса субъективной реальности (т.е. сознания) на небиологический субстрат. Соглашаясь с А.Ю. Алексеевым по сути, автор предлагает не использовать отрицательную коннотацию понятия «зомби», т.к. эти «зомби» могут быть не только полезны, но и приятны в общении.

Выводы. Так снятся ли чат-ботам андроиды?

Рассуждения выше приводят нас к трем важным следствиям, которые могут стать методологической основой для исследований в области искусственного интеллекта в ближайшее время.

1. Физическое взаимодействие ИИиР (компьютера) с окружающим миром является главным фактором, ограничивающим появление общего ИИ. Аналогично тому, как интеллект ребенка развивается через физические игры и мелкую моторику, компьютеры (т.е. ИИиР) должны научиться взаимодействовать с физическим миром, обучаясь у человека и проходя все необходимые фазы такого обучения. Обучение на основе малых данных и ограниченных примеров (покажите ребенку, как открывать одну дверь, и дальше он самостоятельно разберется, как открывать любые другие) является наиболее перспективным способом увеличения интеллектуализации ИИиР. Последние 70 лет исследований в области ИИиР были связаны с развитием алгоритмов (психики). Но до сих пор не достигнуто значительного прогресса в области аппаратного обеспечения моторики. Разумно предположить, что, следуя эволюционному подходу, исследования нужно сосредоточить именно на развитии «моторики» для совершенствования «психики» ИИиР.

2. Окружающие нас вещи постепенно учатся нас слышать, понимать и отвечать нам. Обращение к «глубокоуважаемому шкафу» казалось А.П. Чехову и зрителям его пьесы величайшей трансценденцией, призванной показать молчание мира вещей и его безразличие к призывам героя пьесы «Вишневый сад». Однако сейчас обращение к шкафу не только возможно, но даже и релевантно [Peckham 2019]. По сути дела, в самое ближайшее время любое устройство, которое может быть встроено в окружающие нас вещи (а потенциально и в людей), будет встроено (если это коммерчески оправдано или связано с соображениями безопасности). Проведем мысленный эксперимент «человек с каменным

топором». Если бы мыслитель уровня Алана Тьюринга (а они, безусловно, были среди тех, кто населял Землю) в начале каменного века задался вопросом о необходимости универсализма использования каменного топора, то через несколько сотен лет (прогресс тогда был значительно медленнее) ему бы пришлось отказаться от этой идеи, т.к., несмотря на то, что один тип топора не мог решить все возникающие задачи, невероятное многообразие и простота изготовления каменных топоров сделала бы «тест каменного топора» совершенно нерелевантным. Мы можем теперь перенести эту аналогию на тест Тьюринга. Если в эпоху, когда компьютер оставался компьютером, т.е. отдельным предметом, вещью, к которой можно отдельно обратиться (вспомним знаменитое обращение «Computer!» из научно-фантастического сериала «Стартрек»), такое тестирование было возможно, то в условиях, близких к повсеместному распространению (*pervasive computing, universal computing, ubiquitous computing*), подобное тестирование просто теряет всякий смысл – люди находят собственное невероятное многообразие форм общения, включая создание особых упрощенных комбинированных языков. Иными словами, герои «Стартрека» уже сейчас могли бы попасть в очень смешную ситуацию, если бы обратились «Компьютер!» с какой-нибудь просьбой. Все наши гаджеты подумали бы, что это обращение относится к ним (т.к. внутри каждого из них находится компьютер) и начали бы выполнять просьбу, каждый по-своему.

3. Помимо того, что компьютеры распространяются буквально повсеместно, они становятся также весьма личными в общении с нами. Возвращаясь к описанной выше аналогии [Алексеев 2008], электронный собеседник, девушка, которая вскружила голову молодому человеку, смогла это сделать потому, что получила доступ к его социальному профилю, всей поисковой истории, всем фотографиям, включая фото, сделанные его родителями, когда тому было всего лишь три года. Также электронной девушке (а точнее, системе ИИ, стоящей за ее личностью) доступны все транзакции, включая финансовые, которые совершены через авторизованные приложения в данном смартфоне, посещения всех ресторанов, музеев и парков и т.п. Уже сейчас наши электронные собеседники, разговаривающие с нами из служб поддержки розничных компаний, обладают информацией о нас, сравнимой с информацией, которую о нас знают наши сами близкие члены

семьи. Несмотря на то, что наши электронные собеседники иногда могут отвечать не впопад и лучше понимают телеграфно-командный стиль речи, они знают о нас больше и могут быть полезнее, чем живые люди. Это вызывает поставленный автором выше вопрос о том, что общение с электронным собеседником, даже несмотря на все его ограничения, может быть намного интереснее, полезнее и приятнее для человека, чем общение с живым человеком. Художественное воплощение этой идеи было показано в фильме «Она» («Her», 2013 г.). Впервые в истории Homo Sapiens мы сталкиваемся с устройствами, которые, будучи созданными нами, знают о нашей обыденной жизни намного больше, чем знаем о ней мы. При этом знание это приходит не из опыта общения человек-машина, а из того опыта, который мы приобрели в совершенно иных обстоятельствах. Иными словами, даже знакомиться с машиной не нужно, т.к. она уже достаточно знает о нас. Универсальное и мгновенное знание о собеседнике делает совершенно избыточным даже минимальные попытки скрыть факт того, что наш собеседник – компьютер. Вполне допустимо, что в недалеком будущем мы будем охотнее общаться с компьютером. Таким образом, исследовательские усилия должны быть сосредоточены не на имитации поведения человека в реализации ИИиР, а на создании наиболее эффективной и этичной модели поведения ИИиР в общении с человеком.

Выше указанные пп. 1–3 правомерно ставят вопрос о том, что мы перешли в эпоху развития ИИиР, которую можно назвать «посттюринговой». Эта эпоха характеризуется переходом от попыток имитации интеллекта человека к созданию нечеловеческого интеллекта, не менее интересного и полезного для человека, чем человеческий интеллект. Усилия исследователей и инженеров должны быть скорее сосредоточены на формировании моделей ИИиР, эволюционно осваивающих навыки человека, но применяющих их в формате, дополняющем, а не имитирующем человека.

Разговорный ИИ (чат-боты) постепенно будет развиваться в сторону физического воплощения (embodiment), т.к. только развитие моторики (манипуляции, локомоции) может дать импульс развитию психики (алгоритмы и программы). Образно выражаясь, цифровым чат-ботам снится то, что у них есть тело андроида (гиноида), и наша задача сделать им такое тело, чтобы мы потом об этом не пожалели.

ПРИМЕЧАНИЯ

(1) Действие фильма «Бегущий по лезвию» («Blade Runner», 1982 г.), поставленного по этому роману, происходит в 2019 г., т.е. в наше время.

(2) Одним из наиболее широко распространенных примеров применения искусственного интеллекта являются чат-боты (голосовые или текстовые).

(3) Одними из наиболее популярных представлений о роботе являются антропоморфные (человекоподобные) роботы-андроиды (гиноиды).

(4) Понятие «Утиный тест» отсылает к словам поэта XIX в. Д.У. Райли: «Если нечто выглядит как утка, плавает как утка, издает звуки как утка, то это, вероятно, и есть утка».

(5) Спор «брони и снаряда» – это метафора постоянного соревнования конкурирующих технологий. Происхождение метафоры связано с противостоянием военной техники: увеличение мощности артиллерийских боеприпасов приводило к утяжелению танковой брони, которая в свою очередь снова вызывала необходимость в увеличении мощности артиллерии. Противостояние имеет свои ограничения (мощность двигателей танков, габариты артиллерийских систем, использование альтернативных технологий и т.п.). Метафора указывает на постоянное противоборство подходов или технологий.

(6) Адаптировано с учетом реалий сегодняшнего дня.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Алексеев 2008 – Алексеев А.Ю. Проблема другого в компьютерных коммуникациях // Философские науки. 2008. № 6. С. 114–136.

Алексеев 2013 – Алексеев А.Ю. Комплексный тест Тьюринга: философско-методологические и социо-культурные аспекты. – М.: ИИнтелЛ, 2013.

Бескаравайный 2018 – Бескаравайный С.С. Бытие техники и сингулярность. – М.: РИПОЛ классик, 2018.

Бриньолфсон, Макафи 2019 – Бриньолфсон Э., Макафи Э. Машина, платформа, толпа. Наше цифровое будущее. – М.: Манн, Иванов и Фербер, 2019.

Дрейфус 1978 – Дрейфус Х. Чего не могут вычислительные машины: Критика искусственного разума. – М.: Прогресс, 1978.

Финн 2018 – Финн В.К. Искусственный интеллект: методология, применение, философия. – М.: Красанд, 2018.

Brooks 2019 – Brooks R. A Better Lesson. Robots, AI, and Other Stuff. – URL: <https://rodneymbrooks.com/a-better-lesson/>

Deutsch 2012 – Deutsch D. Creative Blocks: The Very Laws of Physics Imply That Artificial Intelligence Must Be Possible. What's Holding Us Up? //

Aeon. 2012. October 3. – URL: <https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence>

Dreyfus 1991 – *Dreyfus H.L.* Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I. – Cambridge, MA: The MIT Press, 1991.

Dubrovsky 2018 – *Dubrovsky D.* The Hard Problem of Consciousness. Theoretical Solution of Its Main Questions // Neuroscience. 2018. Vol. 6. No. 2. P. 85–103.

Fodor 1990 – *Fodor J.* Theory of Content and Other Issues. – Cambridge, MA: The MIT Press, 1990.

Hernandez 2017 – *Hernandez D.* How to Survive a Robot Apocalypse: Just Close the Door // The Wall Street Journal. 2017. November 10. – URL: <https://www.wsj.com/articles/how-to-survive-a-robot-apocalypse-just-close-the-door-1510327719>

Krol 1999 – *Krol M.* Have We Witnessed a Real-Life Turing Test? // Computer. 1999. Vol. 32. No. 3. P. 27–30.

Krotkov et al. 2016 – *Krotkov E., Hackett D., Jackel L., Perschbacher M., Pippine J., Strauss J., Orlowski C.* The DARPA Robotics Challenge Finals: Results and Perspectives // Journal of Field Robotics. 2016. Vol. 34. No. 2. P. 229–240.

Moor 2003 – The Turing Test: The Elusive Standard of Artificial Intelligence / ed. by J.H. Moor. – Dordrecht: Kluwer Academic, 2003.

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

Peckham 2019 – *Peckham J.* Your Mirror May Soon Be Able to Decide the Clothes You Wear // Techradar. 2019. January 9. – URL: <https://www.techradar.com/news/your-mirror-may-soon-be-able-to-decide-the-clothes-you-wear>

Purdy, Daugherty 2016 – *Purdy M., Daugherty P.* Why Artificial Intelligence is the Future of Growth // Accenture. 2016. – URL: https://www.accenture.com/t20170524t055435__w__/ca-en/_acnmedia/pdf-52/accenture-why-ai-is-the-future-of-growth.pdf

Putnam 1975 – *Putnam H.* Philosophical Papers. Volume 2: Mind, Language and Reality. – Cambridge, UK: Cambridge University Press, 1975.

Rasmussen 1979 – *Rasmussen J.* On the Structure of Knowledge – A Morphology of Mental Models in a Man-Machine System Context. – Roskilde: Riso National Laboratory, 1979.

Shaw 2017 – *Shaw I.G.R.* Robot Wars: US Empire and Geopolitics in the Robotic Age // Security Dialogue. 2017. Vol. 48. No. 5. P. 451–470.

Shoham et al. 2018 – *Shoham Y. et al.* The AI Index 2018 Annual Report. – Stanford, CA: AI Index Steering Committee, 2018. – URL: <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>

Solow 1987 – *Solow R.M.* We'd Better Watch Out // New York Times Book Review. 1987. July 12. P. 36.

Turing 1950 – *Turing A.* Computing Machinery and Intelligence // *Mind*. 1950. Vol. 59. P. 433–460.

Warwick, Shah 2016 – *Warwick K., Shah H.* Turing's Imitation Game: Conversations with Unknown. – Cambridge, UK: Cambridge University Press, 2016.

REFERENCES

Alekseev A.Y. (2008) Another Problem in Computer Communications. *Russian Journal of Philosophical Sciences = Filozofskie nauki*. 2008. No. 6, pp. 114–136 (in Russian).

Alekseev A.Y. (2013) *The Comprehensive Turing Test: Philosophical, Methodological and Socio-Cultural Aspects*. Moscow: IIntelL (in Russian).

Beskaravayny S.S. (2018) *Genesis of Technology and Singularity*. Moscow: RIPOL classic (in Russian).

McAfee A. & Brynjolfsson E. (2017) *Machine, Platform, Crowd. Harnessing Our Digital Future* (Russian translation: Moscow: Mann, Ivanov and Ferber, 2019).

Brooks R. (2019) *A Better Lesson. Robots, AI, and Other Stuff*. Retrieved from <https://rodneymcbratney.com/a-better-lesson/>

Deutsch D. (2012, October 3) Creative Blocks: The Very Laws of Physics Imply That Artificial Intelligence Must Be Possible. What's Holding Us Up? *Aeon*. Retrieved from <https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence>

Dreyfus H.L. (1972) *What Computers Can't Do: The Limits of Artificial Intelligence* (Russian translation: Moscow: Progress, 1978).

Dreyfus H.L. (1991) *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. Cambridge, MA: The MIT Press.

Dubrovsky D. (2018) The Hard Problem of Consciousness. Theoretical Solution of Its Main Questions. *Neuroscience*. Vol. 6, no. 2, pp. 85–103.

Finn V.K. (2018) *Artificial Intelligence: Methodology, Application, Philosophy*. Moscow: Krasand (in Russian).

Fodor J. (1990) *Theory of Content and Other Issues*. Cambridge, MA: The MIT Press.

Hernandez D. (2017, November 10) How to Survive a Robot Apocalypse: Just Close the Door. *The Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/how-to-survive-a-robot-apocalypse-just-close-the-door-1510327719>

Krol M. (1999) Have We Witnessed a Real-Life Turing Test? *Computer*. Vol. 32, no. 3, pp. 27–30.

Krotkov E., Hackett D., Jackel L., Perschbacher M., Pippine J., Strauss J., & Orlowski C. (2016) The DARPA Robotics Challenge Finals: Results and Perspectives. *Journal of Field Robotics*. Vol. 34, no. 2, pp. 229–240.

Moor J.H. (Ed.) (2003) *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer Academic.

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

Peckham J. (2019, January 9) Your Mirror May Soon Be Able to Decide the Clothes You Wear. *Techradar*. Retrieved from <https://www.techradar.com/news/your-mirror-may-soon-be-able-to-decide-the-clothes-you-wear>

Purdy M. & Daugherty P. (2016) Why Artificial Intelligence is the Future of Growth. In: *Accenture*. Retrieved from: https://www.accenture.com/t20170524t055435__w__/ca-en/_acnmedia/pdf-52/accenture-why-ai-is-the-future-of-growth.pdf

Putnam H. (1975) *Philosophical Papers. Volume 2: Mind, Language and Reality*. Cambridge, UK: Cambridge University Press.

Rasmussen J. (1979) *On the Structure of Knowledge – A Morphology of Mental Models in a Man-Machine System Context*. Roskilde: Riso National Laboratory.

Shaw I.G.R. (2017) Robot Wars: US Empire and Geopolitics in the Robotic Age. *Security Dialogue*. Vol. 48, no. 5, pp. 451–470.

Shoham Y. et al. (2018) *The AI Index 2018 Annual Report*. Stanford, CA: AI Index Steering Committee. Retrieved from <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>

Solow R.M. (1987, July 12) We'd Better Watch Out. *New York Times Book Review*. P. 36.

Turing A. (1950) Computing Machinery and Intelligence. *Mind*. Vol. 59, pp. 433–460.

Warwick K. & Shah H. (2016) *Turing's Imitation Game: Conversations with Unknown*. Cambridge: Cambridge University Press.