

ФИЛОСОФИЯ

Горбачева Анна Геннадьевна

ассистент

ФГБОУ ВО «Новосибирский государственный
университет экономики и управления»

г. Новосибирск, Россия

Мельчукова Любовь Валерьевна

ФГБОУ ВО «Новосибирский государственный
университет экономики и управления»

старший преподаватель

г. Новосибирск, Россия

Евстратенко Елена Сергеевна

ассистент

ФГБОУ ВО «Новосибирский государственный
университет экономики и управления»

г. Новосибирск, Россия

ФИЛОСОФСКИЙ АНАЛИЗ ПРОБЛЕМЫ СОЗДАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация: в статье мысленные эксперименты Тьюринга и Серля рассмотрены с учетом стремительного развития информационных технологий. Введены такие понятия, как «контекстный искусственный интеллект», «искусственный интеллект определенного уровня» и «абсолютный искусственный интеллект». Авторами рассмотрены примеры, демонстрирующие, что машина сможет пройти тест Тьюринга благодаря изменению взаимодействия между человеком и машиной и происходящим процессом снижения уровня человеческого интеллекта.

Ключевые слова: искусственный интеллект, тест Тьюринга, «Китайская комната».

Быстрое развитие информационных технологий приближает создание искусственного интеллекта. Время, когда появится искусственный интеллект, будет называться сингулярностью. Данное название предложил Джон фон Нейман – один из основателей цифровых вычислений [2], а его идею продолжили Вернор Винж и Рей Курцвейл, предположившие, что сингулярность появится в третьем десятилетии XXI века [2]. Это время, когда компьютеры и роботы будут способны воспроизводить самих себя и при каждом новом воспроизводстве будут становиться все совершеннее и умнее. На первых этапах человек сможет контролировать этот процесс, но в один прекрасный момент скорость усовершенствования приведет к созданию «суперинтеллекта», который захватит власть на Земле. Важно понимать, что сингулярность влечет физическое уничтожение людей и загрузку их сознания в компьютер, однако при этом человек будет способен осознавать себя [16]. Данная антигуманистическая версия заставляет остановиться и задуматься, но отвернуться от неё уже невозможно.

В недалеком прошлом, идеи по созданию искусственного интеллекта были скрыты в лабораториях, но со временем развитием технокультуры данный вопрос, стал открываться миру.

Возможность создания искусственного интеллекта всегда вызывает устойчивый интерес как у специалистов в области информатики, так и у философов, пытающихся ответить на вопрос «Может ли машина мыслить?». Основоположником данного направления исследований является А. Тьюринг, предположивший, что через 50 лет после его публикации (в 2000 году) мыслящая машина будет создана [14]. Однако, несмотря на усилия ученых, такая машина на сегодняшний день не создана. Более того, Д. Серль предложил опровержение данной гипотезы в виде концепции так называемой «Китайской комнаты» [15]. Он утверждал, что машина способна пройти тест Тьюринга даже, не «умая» мыслить, т. е. не понимая семантики языка. Ей достаточно уметь оперировать с синтаксисом: поверхностными грамматическими структурами и правилами их построения.

Целью настоящей работы является анализ мысленных экспериментов Тьюринга и Серля с учетом стремительного развития информационных технологий, поспособствовавших передаче целого ряда человеческих функций машинам. Основная позиция, с которой мы рассмотрим тест Тьюринга – это его практическая значимость. Нас будут интересовать именно то как искусственный интеллект взаимодействует с внешней средой, а не наличие или отсутствие у нее сознания и понимания вопросов и ответов.

Тьюринг представил свою идею в виде мысленного эксперимента, который он основал на известной в викторианское время салонной игре. Предполагалось, что мужчина и женщина укрывались, а арбитр (собеседник-исследователь) должен был выяснить, кто, есть кто, основываясь лишь только на передаваемых записках. Тьюринг заменил женщину компьютером. Арбитр (собеседник-исследователь) взаимодействуя с машиной (претендующей на обладание искусственным интеллектом) или человеком, должен определить, с кем из них он взаимодействует. Согласно гипотезе Тьюринга, машина, которую арбитр (собеседник-исследователь) не сможет отличить от человека, обладает искусственным интеллектом.

В известной книге «Вы не гаджет» Д. Ланир написал, что тест Тьюринга работает в обе стороны. Люди не могут определить, что машина стала умнее, просто потому, что интеллектуальные способности самих людей становятся ниже [9].

Серль предложил следующий эксперимент. В некоторую комнату помещают человека, не знающего китайского языка, но обладающего инструкцией с описанием правил составления ответов на вопросы, записанные на китайском языке. Серль утверждал, что такой человек, руководствуясь инструкцией, сможет составлять адекватные ответы, не понимая смысла ни вопросов, ни ответов.

Несмотря на критику теста Тьюринга со стороны Серля и его последователей, интерес к проблеме создания машины, успешно проходящей данный тест, по-прежнему высок. Создание такой машины оказывается интересным уже само

по себе, несмотря на противоречия относительно того, можно ли считать успешное прохождение теста свидетельством наличия интеллекта у машины [8]. Общение с компьютерным персонажем, созданным программой искусственного интеллекта, могут заставить задуматься о степени человеческой деградации, которая отвечает за степень реалистичности общения с виртуальным персонажем. На сегодняшний день происходит снижение интеллектуальных способностей человека, и одновременно на этом фоне интеллектуальные устройства выглядят весьма разумно [3; 6].

Этому способствует множество причин. Например, использование интеллектуальных протезов или современная система обучения, требующая от преподавателей натаскивание обучающихся на прохождение тестов, чтобы студенты хорошо выглядели с точки зрения алгоритмов программы. В сети превалирует общение образами [7], а в повседневной жизни – жизненный аутсорсинг [11]. Поисковики стали настолько умными, что поисковиков, что уже предлагают готовые варианты запроса. Причем здесь непонятно, действительно ли поисковик знал, что нужно человеку, или человеку просто не хочется формулировать свой запрос [9]. Ланир пишет, что восприятия человека будет меняться в результате взаимодействия с новыми технологиями: чтобы человек рассматривал разум машины как реальный, требуется ослабить собственную связь с реальностью [9]. Ослабление связи с реальностью происходит из-за развития технологий: уже сейчас есть «умные дома» где компьютер регулирует температуру дома, независимое существование домашней техники. Все перечисленное безусловно предоставляет все большую свободу для человека и при этом все более отрывает его от реального физического мира. Для многих виртуальный мир становится более реалистичным, чем реальный физический мир.

Данную ситуацию еще усугубляет современное образование, где применяются новые методы обучения, учитывающих особенности мышления современного человека [1], что приводит к изменению мышления [4]. Стоит также привести слова, бывшего министра образования Российской Федерации С. Фурсенко

который высказал идею: «недостатком советской системы образования была попытка формировать человека-творца, а сейчас задача заключается в том, чтобы взрастить квалифицированного потребителя, способного квалифицированно пользоваться результатами творчества других» [12].

Все выше перечисленное, нас заставляет по-новому рассмотреть концепции Серля и Тьюринга. Прежде всего нас будут интересовать участвовавшие взаимодействия человека и машины, человека и тех интеллектуальных устройств, которыми человек ежедневно пользуется. Напомним, что в тесте Тьюринга испытуемым объектом являлась машина, которая «пыталась» успешно пройти тест. Тьюринг считал, создание такой машины делом весьма недалекого будущего и уже к двухтысячному году считал, что тест будет пройден.

Одним из следствия известного закона Мура можно считать экспоненциальный рост производительности процессора ЭВМ, точнее наблюдается удвоение производительности процессора каждые восемнадцать месяцев, другими словами, компьютеры в два раза становятся мощнее. К тому же, для того чтобы машина смогла пройти тест Тьюринга, программисты все более совершенствуют программы. При этом не учитывается, что если арбитром (собеседником-исследователем) выбрать человека с низким уровнем интеллекта, не умеющего задавать каверзные, сложные вопросы, или человека в состоянии измененного сознания, например, под воздействием алкоголя или наркотиков, то он скорее всего не сможет отличить человека от машины даже с относительно простой программой. Следовательно, машина, будет способна проходить тест не только когда ее интеллект уподобится человеческому, но и когда арбитр (собеседник-исследователь) будет обладать низким уровнем интеллекта. Из этого можно сделать вывод, что постановка теста Тьюринга не совсем корректна, особенно в наше время, когда современный человек все больше передает свои интеллектуальные функции техническим устройствам. Происходит размывание – где здесь человек, а где машина? В работе мы предложим понятия «интеллекта определенного уровня, контекстного искусственного интеллекта и «абсолютного» искусственного интеллекта».

Тьюринг один из первых задал вопрос: «Может ли машина мыслить?». Через много лет спустя Серль его уточнил: «Могут ли машины понимать и иметь осознанные мысли?». Но можно ли таким образом задавать вопрос? Уже сегодня существуют примеры того, что тест Тьюринга был практически пройден, за счет *степени персонификации*. Например, в игре в шахматы или покупку через интерфейс, которую мы опишем ниже. Далее мы приведем понятие контекстного искусственного интеллекта и докажем, что искусственный интеллект уже на этом этапе создан. Но для начала приведем и другие факты наличия искусственного интеллекта. Например, в 1997 году программа Deep Blue в чемпионате мира по шахматам обыграла Гарри Каспарова [13]. Победа компьютера в этом соревновании вызвала в страх в обществе, так как игра в шахматы всё это время считалась наивысшим видом разума, которым обладает исключительно человек, но одержанная победа, показала, что компьютер может играть лучше.

Это не первый случай, когда компьютер опередил человека, в 1982 году по мнению журнала Time явлением года, был назван персональный компьютер – первый лауреат, неодушевленный предмет. Традиционно этот журнал ориентировался на поиски человека года, и провозглашение 1982 года компьютера показало, что в сферу живого, в сферу чисто человеческих отношений вторгается машина. Мало того, что она вторгается, она ещё занимает лидирующую роль, опережая человека.

Следует так же заметить, что мы каждый день сталкиваемся с искусственным интеллектом, когда взаимодействуем с банкоматом, когда компьютер решат за нас определенные задачи и т. д. Мы полагаем, «умение» машины проходить успешно тест Тьюринга важнее, чем понимать семантику своих действий. Искусственный интеллект разрабатывается не только для теоретического интереса, но главной его целью является принесение пользы для общества. В.П. Литвинов, считает, что Тьюринг видел предназначение искусственного интеллекта для решения «конструктивных-технических задач» [10].

В данной работе мы предложим понятие контекстного искусственного интеллекта и обоснуем, что он уже создан и имеет важное практическое значение.

Актуальные направления научных исследований: от теории к практике

Мы будем считать, что машина обладает *контекстным искусственным интеллектом*, если она успешно проходит тест Тьюринга в конкретном контексте и непредвзятом исследователе, но может не пройти тест, если хотя бы одно из этих условий не выполнено. Под контекстом будем понимать комбинацию цели взаимодействия и правил (интерфейса) взаимодействия.

В мысленном эксперименте Тьюринг не предполагал никаких ограничений на вопросы, которые может задавать арбитр (собеседник-исследователь). Вопросы могли быть самыми разными и неожиданными. Это привело к тому, что ни одна машина его не может пройти. Главная проблема, на наш взгляд, заключается в неточной постановке задачи. Чтобы хоть как-то решить проблему искусственного интеллекта, необходимо поставить некоторые разумные ограничения на условия, в которых будет проводиться тестирование.

Рассмотрим следующий пример: приобретение товаров через автоматизированные терминалы. Это самый яркий пример контекстного искусственного интеллекта. В этом примере совершенно ясна цель взаимодействия – покупка товара, и определен интерфейс, через который происходит взаимодействие пользователя с компьютером. Человек взаимодействует с терминалом, нажимая нужные кнопки и передавая денежные средства. Если человек будет приобретать товар у продавца через точно такой же интерфейс, то отличить машину от человека он не сможет.

Теперь рассмотрим выход за рамки установленного контекста. Представим, что главная цель покупателя состоит в выяснении того, машина это или человек. Тогда в данном случае исследователь предвзят, и его цель не соответствует цели создания данного терминала. Исследователь, который поставил цель выяснить, машина это или человек, может, например, взломать терминал. В этом случае покупатель нарушил цель создания контекста. Если покупатель вздумает поговорить с терминалом, как часто делают в магазинах постоянные клиенты, то он также выйдет за рамки контекста.

Второе понятие, которое мы введем – это *искусственный интеллект определенного уровня*. К настоящему моменту нет строгих доказательств наличия различного уровня интеллекта у разных людей, но существуют следующие уровни измерения интеллекта, например, оценки в образовательных учреждениях или тест IQ. Точно доверять перечисленным методам исследования сложно, тем не менее, можно выдвинуть следующее предположение:

1. Уровень интеллекта у разных людей разный.
2. Способ измерения уровня интеллекта должен существовать, хотя он еще и не открыт.

В рамках этих предположений можно ввести способ измерения уровня искусственного интеллекта. Искусственный интеллект уровня *X* – это такая система, которая проходит тест Тьюринга, если исследователь имеет интеллект уровня *X*. Искусственный интеллект определенного уровня можно использоваться как тест человеческого интеллекта – человек имеет такой уровень интеллекта, какой уровень искусственного интеллекта он в состоянии распознать. Следовательно, в этом заключается его практическая значимость.

Под абсолютным искусственным интеллектом мы будем иметь в виду такой интеллект, который невозможно отличить от человеческого ни при каких обстоятельствах. Абсолютный искусственный интеллект – способен успешно пройти тест Тьюринга при любом арбитре (собеседнике-исследователе), но все-таки будет оставаться неизвестным, будет ли он по Серлю понимать семантику вопросов и своих ответов. Мы будем считать, что абсолютный искусственный интеллект не будет отличаться от человеческого.

Абсолютный искусственный интеллект будет способен синтезировать знания и направлять их в пользу выживания и приобретения власти. Основным тестом на абсолютный искусственный интеллект будет являться тот факт, что машина будет затруднять человеку «нажимать кнопку», для того, чтобы не дать себя отключить. Уже сейчас во многих автоматизированных системах человек не может отключить машину, например, на атомных станциях, осознавая, что это приведет к катастрофическим последствиям, или же сами программисты создают

определенный уровень защиты от выключения. Когда мы говорим об искусственном интеллекте, то мы не это имеем ввиду. В этих примерах человек не нажимает кнопку отключения по своей инициативе: либо просто осознавая масштабы последствия своего действия, или разработчики спроектировали систему таким образом, чтобы ее было трудно отключить. Мы имеем ввиду, то, что, даже если человек попытается отключить компьютер, он не сможет этого сделать.

Заключение. В настоящей работе приведены аргументы в пользу значимости всестороннего анализа теста Тьюринга. Введены понятия *контекстного искусственного интеллекта*, *искусственного интеллекта определенного уровня* и *абсолютного искусственного интеллекта*. Приведены аргументы в пользу того, что машина сможет пройти тест Тьюринга не только благодаря своему совершенствованию, но и благодаря изменению процесса взаимодействия человека и машины вкупе с возможным снижением человеческого интеллекта. Спрогнозированы возможные последствия создания абсолютного искусственного интеллекта и предложен тест для него.

Список литературы

1. Алексеева И.Ю. Информационная компетентность, естественный интеллект и НБИКС-революция / И.Ю. Алексеева // Информационное общество. – 2012. – №5. – С. 9–15.
2. Виндж В. Технологическая Сингулярность – Компьютерра ONLINE. – [Электронный ресурс]. – Режим доступа: <http://www.computerra.ru/think/35636/>
3. Горбачева А.Г. Человеческий интеллект: возможные изменения под влиянием информационных технологий и высокотехнологичных устройств / А.Г. Горбачева // Идеи и идеалы. – 2014. – №1. – Т. 2. – С. 135–142.
4. Горбачева А.Г. Философский анализ изменения мыслительного процесса людей в постиндустриальном обществе и футурологический образ «человека ассоциирующего» / А.Г. Горбачева // Наука о человеке: гуманитарные исследования. – 2015. – №1 (19). – С. 30–35.

5. Горбачева А.Г. Обмен визуальной информацией и короткими сообщениями как современный вид сетевых коммуникаций / А.Г. Горбачева // Праксема. Проблемы визуальной семиотики. – 2015. – №1 (3). – С. 133–139.

6. Горбачева А.Г. Философское осмысление влияния НБИКС-технологий на интеллект человека / А.Г. Горбачева // Вестник Омского университета. – 2014. – №3 (73). – С. 93–95.

7. Горбачева А.Г. Особенности письменных жанров в эпоху сетевых технологий (языковые идентичности и их метаморфозы) / А.Г. Горбачева // Идеи и идеалы. – 2014. – Т. 2. – №3 (21). – С. 46–54.

8. Горбачева А.Г. Тест Тьюринга: взгляд через призму современных компьютерных и сетевых технологий / А.Г. Горбачева // Вестник НГУЭУ. – 2014. – №4. – С. 322–330.

9. Ланир Д. Вы не гаджет Манифест / Д. Ланир. – М., 2011.

10. Литвинов В.П. Актуальность задачи Тьюринга (Can machines think?) / В.П. Литвинов // Философские проблемы информационных технологий и киберпространства. – 2012. – №1. – С. 93–100.

11. Смирнов С.А. Фармацевтика антропологических трендов. Антропологический форсайт / С.А. Смирнов // Вестник НГУЭУ. – 2012. – №1. – С. 88–104.

12. Правда-TV.ru. [Электронный ресурс]. – Режим доступа: <http://www.pravda-tv.ru/2012/04/19/14237>

13. ПостНаука [Электронный ресурс]. – Режим доступа: <http://postnauka.ru/video/47550>

14. Turing 1950 – Turing A. Computing machinery and intelligence // Mind. – 1950. – №59. – P. 433–460.

15. Searle 1980 – Searle J. Minds, brains, and programs // The behavioral and brain sciences. – 1980. – №3. – P. 417–457.

16. Kurzweil R. The Singularity Is Near: When Humans Transcend Biology. – Penguin, 2006.