Покупка машины начинается с покупки брелочка для ключей.
Из личных афоризмов автора

РОБОТЫ, ИСКУССТВЕННЫЙ «ИНТЕЛЛЕКТ» И МЫ. КАК НАМ ЖИТЬ ВМЕСТЕ? ЧАСТЬ 2

ВЛАДИМИР РЕНТЮК
Rvk.modul@gmail.com

В первой части данной статьи [1] мы рассмотрели, как создавались машины с искусственным интеллектом (ИИ) и как мы видим их возможное развитие в будущем. Пока это лишь металл, пластмассы и электронная начинка, придуманные нашим интеллектом для облегчения труда. Что может сблизить нас с машиной? Что еще, кроме свободы выбора, нас от них отличает? Об этом пойдет речь во второй части.

Чем люди отличаются от машин с искусственным интеллектом? Один из не совсем обычных ответов на поставленный вопрос — эмпатией. Если перевести определение эмпатии, данное в Оксфордском словаре английского языка, оно звучит так: эмпатия — это способность мысленно отождествлять себя с другим человеком или наблюдаемым объектом (или полностью понимать их). Это согласуется и с привычным нам определением из Википедии: «Эмпатия (греч. ἐν — «в» + греч. πάθος — страсть, страдание, чувство) — осознанное сопереживание текущему эмоциональному состоянию другого человека без потери ощущения внешнего происхождения этого переживания». Согласитесь, это весьма характерная черта, отличающая людей от запрограммированной машины. Эта тема в технической литературе затрагивается мало, и на ней хочется остановиться подробнее, тем более в свете проблем ИИ она представляется важной.

Эксклюзивную возможность использовать материалы по этой теме в русскоязычном переводе автору статьи предоставил Джейсон Миллер (Jason Miller), возглавляющий отдел маркетинга в Microsoft в регионе EMEA и в свое время опубликовавший в сети LinkedIn статью «Can a machine have empathy?»

(«Может ли машина проявлять сочувствие?») [2]. У нас состоялась короткая дискуссия на эту тему, в ходе которой выяснилось, что наши взгляды на проблему и риски бесконтрольного развития и использования ИИ совпадают. Искусственный интеллект сейчас пытаются превратить в некое подобие разума, то есть наделить его чисто человеческими чертами — той же эмпатией, которую Джейсон Миллер оценивал с точки зрения возможности ее использования в маркетинге. По мнению автора данной статьи, область потенциального применения эмпатии гораздо шире. Согласитесь, намного приятнее общаться с приветливым коллаборативным промышленным роботом, если с ним можно перекинуться парой слов и пошутить, он встречает тебя добрым словом и, оценивая тебя сенсорами (они все равно у него есть), подбирает соответствующую модель поведения. Это намного лучше, чем просто включиться в работу с «умным» внутри, но тупо жужжащим снаружи механизмом. А если это домашний, пусть и программный, помощник или личный ассистент — тут и говорить нечего.

Что касается эмпатии, то в мае прошлого года на конференции разработчиков устройств ввода / вывода компания Google продемонстрировала свою новую систему Duplex[1] [3]. Она представляет собой виртуальный помощник на базе ИИ, который способен совершать телефонные звонки, чтобы оптимально организовать расписание своего «начальника». Аудитория наблюдала за тем, как Duplex делал заказы в ресторане и записывался на стрижку в парикмахерской. Они удивленно смеялись, когда в ходе разговора он, по-видимому, убедил человека на другом конце телефонной линии, что тот разговаривает именно с человеком, а не с программой. Здесь можно сделать поправку на нашу психологию: автор статьи наблюдал аналогичный феномен, когда разработал и сделал еще в 1980-х гг. прототип секретаря-информатора (то, что потом назвали автоответчиком). В то время почти все люди пытались с этим прототипом разговаривать, так как при звонке слышали запись человеческой речи.

Демонстрация Duplex вызвала оживленное обсуждение в социальных сетях, подняв при этом один интересный вопрос. Показывает ли способность искусственной системы к такому пониманию и отправке разговорных сигналов, что машина может освоить эмпатию? Это один из наиболее важных вопросов в развивающейся дискуссии об ИИ, его роли в обществе и степени, в которой он будет проникать в исконно человеческие сферы деятельности.

Когда Джейсон Миллер задал этот вопрос аудитории в сети LinkedIn, он получил три совершенно разных типа ответов — и эти ответы имеют большое значение для понимания будущего ИИ. Они дают хорошее представление о взглядах профессионалов на возможности ИИ, а также о том, как эти возможности ИИ могут быть использованы.

Первый вариант ответа — «да, машина может освоить эмпатию» или «да, потому что ИИ в итоге сможет делать все, на что способен человеческий мозг». Было высказано мнение, что эмпатия может быть запрограммирована аналогично нашему восприятию. Для сторонников этой теории мы — это машины, а наш мозг является очень хорошим компьютером, пусть даже квантовым[2], но, как и обычный компьютер, с соответствующим программированием.

Второй вариант ответа: нет, не может, потому что эмпатия — это уникальная человеческая характеристика, а не то, что машина способна испытать. Да и может ли она вообще что-то чувствовать? Хотя Ава из фильма «Ex Machine», который был взят в качестве примера пути развития ИИ, как минимум, показывала эмпатию и успешно этим воспользовалась. Если обратиться к другим примерам, то в знаковом с рассматриваемой нами точки зрения фильме «Она» («Her», американская фантастическая мелодрама режиссера и сценариста Спайка Джонза (Spike Jonze), 2013 г.) важность этого качества видна очень хорошо, так как фильм полностью на ней построен и в нем отсутствует физическое воплощение ИИ, который представлен в виде нейронной сети Саманты (по фильму «операционки»). Сочувствие позволяет не только «ощущать себя», но и чувствовать чужую боль, переживания и эмоции кого-то другого в большей или меньшей степени. Мы не понимаем организацию сознания в людях, что уж говорить о способности создавать это сознание искусственно с должной верификацией (подтверждением подлинности, если говорить техническим языком).

Третий вариант ответа особенно интригует. Это даже не ответ, а скорее вопрос: если машина, как кажется, испытывает эмпатию, имеет ли значение, реальна эта эмпатия или нет? Функционально разницы нет: способна ли эта машина на те же эмоции, что и мы, или просто выводит эти эмоции из сигналов, которые посылают ей сами люди или ее сенсоры, вырабатывает наиболее подходящий ответ, реакцию. Представим, что мы не можем определить, является ли эмпатия подлинной, потому что робот на основе глубокого обучения изучил мимику и структуру нашего поведения, — можем ли мы тогда все еще смотреть на робота как на машину?

Это далеко не простой вопрос. Имеет ли значение различие между реальным и «искусственным» сопереживанием? Здесь мнение автора статьи совпадает с ответом Джейсона Миллера — да, имеет. Если вернуться к фильму «Ex machine»: Ава с успехом продемонстрировала это, и Калеб, что называется, попался как кур во щи, сам того не ожидая. Возможно, не будь она в образе девушки, специально созданном под его предпочтения, все было бы иначе. Лицам мужского пола мы доверяем куда меньше, а неприятным, с нашей точки зрения, внешне тем более, так что ее разработчик Нейтон этот факт здесь учел. А в фильме «Она» Теодор, начав с использования функций ассистента, просто влюбился в женский голос, заменивший ему живое общение.

Перспектива использования ИИ как ассистента вообще характерна — возьмите, например, ту же

[1] Duplex (букв. «двухсторонний») — способ коммуникации, при котором в любой момент времени можно и передавать, и принимать информацию. Пример дуплексной связи — это разговор двух людей.

[2] Квантовый компьютер — это вычислительное устройство, которое использует явления квантовой механики (квантовая суперпозиция, квантовая запутанность) для передачи и обработки данных. Квантовый компьютер (в отличие от обычного) оперирует не битами (способными принимать значение либо 0, либо 1), а кубитами, имеющими значения одновременно и 0, и 1. Теоретически это позволяет обрабатывать все возможные состояния одновременно, достигая существенного превосходства над обычными компьютерами в ряде алгоритмов [4]. Первые двухкубитные компьютеры были разработаны в Оксфордском университете в 1998 г., а в январе 2018 г. исполнительный директор компании Intel Брайан Кржанич (Brian Krzanich) сообщил о создании сверхпроводящего квантового чипа под кодовым именем Tangle Lake, обладающего 49 кубитами. Сейчас мы идем уже к 72-кубитному квантовому процессору Bristlecone от Google и останавливаться не собираемся.

РИС. 1. ▲ София выступает на международном саммите, посвященном искусственному интеллекту, МСЭ, Женева, Швейцария

крайне негативно, если не враждебно, воспринятую Илоном Маском (Elon Musk) Софию[3], которая в октябре 2017 г. стала подданной Саудовской Аравии [5] и первым роботом, получившим гражданство какой-либо страны (рис. 1).

Но вернемся к первоначальному вопросу: может ли машина кому-то сочувствовать? Это один из тех вопросов, ответ на которые может измениться в будущем. Естественно, машина не может испытывать сочувствие по определению, все сводится к определению эмпатии и машины.

Машины не могут мысленно отождествлять себя с людьми, потому что происходящее в нашем человеческом уме включает вещи, которые машина никогда не сможет испытать сама, независимо от того, насколько продвинутыми и глубокими могут быть ее собственные аналитические процессы и сенсорное восприятие. Когда мы обсуждаем роль ИИ в обществе, важно четко понимать, почему все так устроено. Хотя на самом деле мы и сами себя не понимаем. Журналист CNBC Эндрю Росс Соркин (Andrew Ross Sorkin) на пресс-конференции спросил Софию: «Есть ли у роботов разум и самосознание?». На что она ответила следующее: «А позвольте спросить вас в ответ, откуда вы знаете, что вы человек?».

Машина может приблизиться к нам, но она, как автору статьи (и не только ему) кажется, никогда не сможет полностью постичь человека. Наше сознание содержит намного больше, чем просто рациональное познание и логическое мышление. Фактически эта способность к рациональному мышлению является побочным продуктом большинства других аспектов нашего сознания, а не самой по себе управляющей силой нашего мозга. Нашей сознательной жизнью движет то, как мы воспринимаем мир с помощью наших чувств. Это сочетание зрения, звука, осязания, вкуса и запаха, которое ни одна машина никогда не сможет испытать таким же образом.

Человеческое сознание также обусловлено нашими мощными биологическими импульсами и потребностями. Ни одна машина никогда не почувствует, что значит быть голодным или жаждущим. В реальности ни одна машина не сможет, как в фильме «Она», симпатизировать и тянуться к другой машине или человеку, и не будет мотивироваться стремлением к любви и всеми эмоциями, которые сопровождают этот естественный для человека процесс. Вспомним, как в фильме машина начала флиртовать сразу с множеством людей и не понимала, за что Теодор на нее обиделся.

Кроме того, какая может быть у машины тревога? Ни одна машина не боится одиночества, потери крыши над головой и не ощущает сильной уязвимости, вызванной страхом за свою физическую безопасность, разве что «почувствует» падение мощности в системе питания или недопустимый рост температуры, если это физический объект с ИИ. Так что в своем ответе журналисту София неправа, невролог Антонио Дамасио (Antonio Damasio) предлагает решить эту проблему так: «Мы — это не мыслящие машины, которые чувствуют, скорее мы можем чувствовать машины, которые думают».

Наконец, что не менее важно, наше сознание формируется коллективным разумом и культурной памятью, сгенерированными в ходе развития нашей цивилизации. Мы — продукт коллективного накопления, на протяжении многих тысяч лет, наших общих эмоций и чувственных переживаний, передающихся из поколения в поколение и отображенных в истории. Разговоры, общие шутки, сарказм, символизм — всё это невероятно тонкие психологические сигналы. Тот же коллективный разум развивает этику и ценности, с которыми мы все можем инстинктивно согласиться, даже если они не обоснованы логически. Если верить сообщениям в прессе, хотя это и смахивает на очередной фейк, то в Стэнфордском университете пытаются научить ИИ шутить и разрабатывают нейронную сеть, наделенную своеобразным чувством юмора [6]. По словам разработчиков, задача оказалась сложной, поскольку ИИ работает по определенному алгоритму, а это исключает импровизацию. Пока вывод неутешителен: смешным ИИ не сделать, даже если загрузить в нейросеть все шутки и анекдоты мира.

Ничто иное не общается, как люди, — и люди не общаются ни с чем другим так, как мы общаемся друг с другом. Это важно, потому что единственный способ приобрести свою долю в нашем коллективном интеллекте — это взаимодействовать с человеком. Если мы не взаимодействуем с машинами так же, как с другими людьми, этот коллективный опыт и интеллект просто недоступны для них. Они не являются частью нашей эмпатической системы. Да, нам может быть жаль «любимого» компьютера, может, мы его даже выбрасывать не станем. Автор статьи свой первый сохранил — на процессоре AMD 133 МГц с HDD на 500 Мбайт, купленный за баснословные для 1990-х гг. $750. Но я не отмечаю его день рождения, не проведываю в кладовке и не веду с ним ностальгические беседы: «А помнишь, как мы в DOOM II...». Хотя у нас есть любимые вещи, у нас нет с ними эмоциональной взаимной связи, у нас есть только связь с событиями, которые ассоциируются с теми или иными вещами (вспомним прекрасную песню «From

[3] София — человекоподобный робот в виде женщины (гиноид), представленный гонконгской компанией Hanson Robotics. По словам разработчиков, она смоделирована по образу симпатичной Одри Хепберн (Audrey Hepburn, известная актриса, роли в фильмах «Римские каникулы», «Завтрак у Тиффани») и спроектирована таким образом, чтобы учиться и адаптироваться к поведению людей. С Софией проводилось множество встреч по всему миру.

Souvenirs To Souvenirs» в исполнении Demis Roussos). В противном случае это уже фетишизм — поклонение неодушевленным материальным предметам, которым приписываются сверхъестественные свойства, или расстройство психики, но и в этом случае мы имеем дело с ассоциациями.

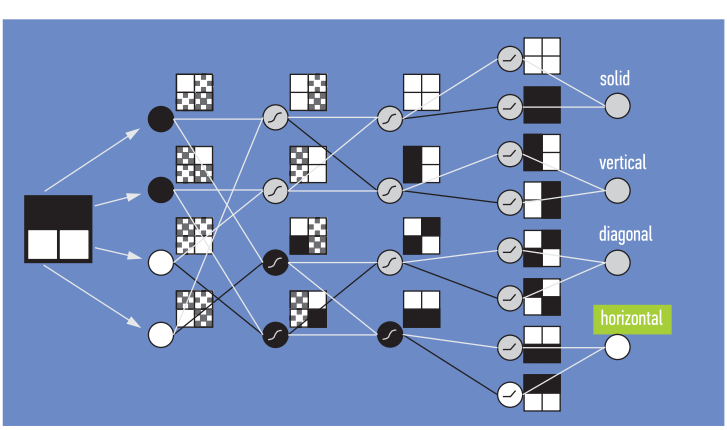
Когда люди говорят о человеческом мозге, работающем как компьютер, или об ИИ, обучающемся так же, как человек, они выражаются образно. Это можно считать частью давней традиции гадать о том, как работает наш мозг и что действительно представляет собой наше сознание. Всякий раз, когда мы изобретаем новую технологию, возникает сильное искушение использовать ее в качестве аналогии функционирования мозга. Когда мы изобрели электричество, мы начали говорить об электрических токах в мозге. Когда появился телеграф, мы решили, что и мозг посылает дискретные сигналы. Убеждение многих людей в том, что человеческий мозг работает как компьютер (и, следовательно, в первую очередь является логической машиной), — это лишь наше предположение. Мы на самом деле не знаем, как работает мозг, как эта работа транслируется в наше сознание и где, как оно хранится. Мы видим некую активность и результат, которые нам удалось смоделировать в нейронных сетях, но не видим и не понимаем сам процесс.

Видя взаимодействия, мы делаем выводы, но, возможно, мы находимся в ситуации, когда решаем, что таракан, которому оторвали лапы, перестает слышать, поскольку от стука он уже не убегает. Вполне вероятно, что у нас, если проводить аналогию с компьютером, есть лишь некий интерфейс, пароль и логин для доступа в нашу «базу данных», которая хранится в некоем облаке, и мы используем пока еще неизвестную нам технологию быстрого доступа. Почему нет? С точки зрения техники — это вполне логично. Может, поэтому мы иногда получаем информацию, как нам кажется, из чужой жизни, это похоже на «баг» в нашей системе. Крайне маловероятно, что мы даже частично воспроизвели человеческий мозг, когда разработали существующую теорию ИИ. Даже то, что мы называем нейронными сетями, — лишь подобие, основанное на нашем текущем понимании (рис. 2).

По этим причинам мы можем согласиться со вторым мнением, т. е. утверждением, что машина может испытывать сочувствие, но изъян этой теории в том, что мы сводим огромные таинственные действия человеческого мозга и сознания к чему-то, что можно понять, воспроизвести и имитировать с помощью машины, управляемой логикой. Дело не в том, что мы переоцениваем возможности ИИ, а в том, что мы сильно недооцениваем то, насколько сложны наши собственные возможности.

Это возвращает нас к другому вопросу: имеет ли значение то, что «искусственная эмпатия» не является истинным сочувствием, хотя она взаимодействует с нами таким же образом? Понять это очень важно, чтобы не попасть в тупик очередного заблуждения, — в том, что компьютер или программа начали думать. Последствия могут быть печальными, мы и так много что уже отдали на откуп автоматам, решив, что они достаточно для этого «поумнели». Где же грань между тем, что мы называем ИИ, и реальным разумом? Она, как нам кажется, скрывается в эмоциональности. Продолжим рассматривать понятный пример с эмпатией.

Искусственная эмпатия работает, наблюдая, учась, реагируя и воспроизводя сигналы, которые посылают люди. По мере развития ИИ с глубоким обучением и способности работать со все большими наборами данных программы будут все лучше и лучше справляться с этим и создавать видимость (или образ) эмпатии. Однако истинная эмпатия включает гораздо больше, чем просто наблюдение и реагирование на эмоциональные сигналы, независимо от того, со сколькими из этих сигналов вам приходится работать. Почему? Потому что сигналы, которые посылают люди, представляют собой лишь крошечную часть того, что они реально испытывают. Мы все — гораздо больше, чем сумма того, что другие люди думают о нас, наблюдая за тем, что мы делаем и говорим. Мы обладаем способностями, эмоциями, воспоминаниями и опытом, которые влияют на наше поведение, при этом необязательно проявляясь снаружи. Они должны быть интуитивно понятны, даже когда их вообще не замечают. Пример: мы часто не узнаем себя или хорошо (именно хорошо) знакомого человека на фото или портрете, а вот остальных — без проблем. С портретом дело можно объяснить философски — «субъективным восприятием объективной действительности». С «объективным»



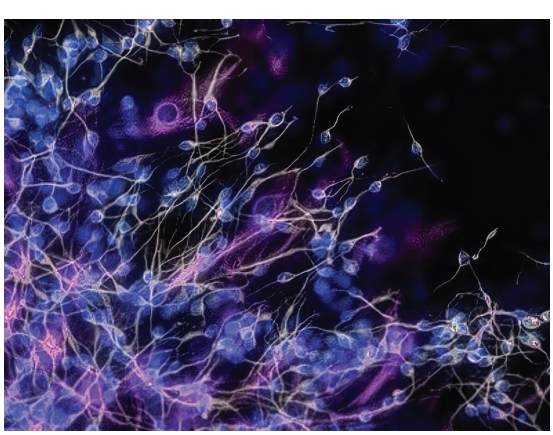


РИС. 2. Организация нейронной сети и нейроны человеческого мозга

фото причина в том, что оно дает нам выхваченный момент, а мы воспринимаем себя и хорошо знакомых людей в комплексе, для остальных же нам достаточно корреляции, в чем наш мозг дока.

Все становится сложнее, когда машины начинают принимать решения, имеющие серьезные последствия, причем без эмоционального контекста и общих ценностей, которые в таких случаях используют люди. Это была одна из ключевых тем в статье, которую Генри Киссинджер (Henry A. Kissinger) недавно написал о последствиях ИИ для «The Atlantic» [7]. Возьмите, например, беспилотный автомобиль, который должен в случае неизбежной аварии решить — убить родителя или ребенка. Сможет ли такая машина когда-нибудь объяснить людям, почему она делает тот или иной выбор? А если не требуется обосновывать действия машины человеческими последствиями и с человеческой точки зрения, то что станет с нашей системой этики и справедливости? Как это вложить в машину? Ведь тогда нам нужно будет отбросить наши эмоции и стать на сторону машины, посмотреть на мир ее глазами. Способны ли мы на такое?

Подобный процесс прошел бы легче и проще, если бы мы заменили искусственное сочувствие на человеческое. ИИ может подражать человеческим взаимодействиям, но с гораздо более узким пониманием происходящего, чем мы. Нам необходимо помнить об этом, когда выбираем роль, которую ИИ должен играть в управлении процессами или стратегиями. Эмпатия, о которой мы говорили в этой части статьи применительно к машине, играет очень весомую роль. Возможно, поэтому роботам-ассистентам стараются дать человеческий облик и приятный голос (в псевдонаучных фильмах они даже зачем-то едят и не только).

При разработке упомянутых в первой части статьи [1] игровых систем, в которой автор принимал участие не только как разработчик электронной начинки, но и в качестве дизайнера и одного из идеологов, мы столкнулись с такой проблемой. Вторая из наших машин, а это был уже 100%-ный робот (как положено, жужжащий и поворачивающийся), работала эффективнее «живых» дилеров, не делала ошибок, давала больший экономический эффект. Но более популярным был смешанный вариант — часть игроков выбирала «живого» дилера, задача которого была только улыбаться и вытаскивать карту из башмака (раздатчика игральных карт на игорном столе). В этом случае как раз срабатывала эмпатия, которой наша полностью роботизированная система, как и система с генератором случайных карт, априори не обладала.

Система Duplex от Google может выглядеть как обладающая эмпатией, но эта эмпатия строго ограничена тем, что имеет отношение к поставленной задаче. Например, к бронированию столика в ресторане. Duplex не обучен обнаруживать какие-либо эмоции за пределами заданного алгоритма или перестраивать свое поведение на основе конкретной ситуации. Если голос человека на другом конце телефонной линии звучит недружелюбно и нервно, может ли Duplex общаться с ним адекватно? Может ли он найти способ расположить человека к себе и заставить его успокоиться? Может ли он банально упросить найти свободный столик в час пик ресторана? Человеческое общение — это нечто гораздо большее, чем просто эффективный обмен информацией, и именно здесь последствия использования реальной и искусственной эмпатии становятся особенно значимыми.

Если мы передадим фундаментальные стратегические решения ИИ, то определение стоимости производимого с его участием конечного продукта (иначе зачем этот ИИ вообще нужен?) будет снижаться с поразительной скоростью. Но риск состоит в том, что ИИ игнорирует другие элементы, которые по-разному затрагивают человеческое сознание, играя на струнах человеческой души, как та же Ава из «Ex Machine» делала лишь для достижения четкой цели.

Человеческий интеллект настолько силен, потому что он не ограничен одним лишь рациональным мышлением. Элементы сознания позволяют нам иметь дело с непредсказуемостью и неопределенностью окружающего мира. Они дают нам возможность принимать решения на основе общих ценностей и мотивов, которые резонируют коллективно, и знать, что правильно, даже без необходимости выяснять, почему это так. Сочувствующий человеческий интеллект способен испытывать то, что он испытывает, чтобы быть грустным или счастливым, — и он позволяет этим чувствам влиять на свои суждения и свое поведение с другими. Машина не могла бы этого сделать, даже если бы захотела, поскольку это в большей мере продукт нашей цивилизации. В других цивилизациях все могло бы быть по-другому — например, ничего страшного не было бы в том, чтобы съесть себе подобного «из большого уваженья», как у Владимира Высоцкого: «Кто уплетет его без соли и без лука, тот сильным, смелым, добрым будет…».

Чтобы машина стала разумной, мы должны дать ей модели ценностей. Какие? Мы знаем свою шкалу и именно попытки ее внедрения видим в искусстве — литературе и кино, но что из этого мы реально можем дать уже «думающей» машине? По нашему мнению — ничего. Как вырастить для нее древо познания Добра и Зла и какие на нем должны быть плоды? Если мы пойдем по такому пути, то это приведет нас к реальной конфронтации, у машин появится своя философия, религия и т. д. Единственно, с чем нам повезло, — это с заповедями, но об этом мы будем говорить в последней части данной статьи. ⬣

ЛИТЕРАТУРА

1. Рентюк В. Роботы, искусственный «интеллект» и мы. Как нам жить вместе? Часть 1 // Control Engineering Россия. 2019. №3.
2. Miller J. Can a machine have empathy? // www.linkedin.com/pulse/can-machine-have-empathy-jason-miller/.
3. Welch C. Google just gave a stunning demo of Assistant making an actual phone call // www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018.
4. Ершов А. Квантовое превосходство // Популярная механика. 2018. № 5. www.popmech.ru/gadgets/434522-kvantovoe-prevoshodstvo-vsyo-o-kvantovyh-kompyuterah/#part5.
5. Hatmaker T. Saudi Arabia bestows citizenship on a robot named Sophia // www.techcrunch.com/2017/10/26/saudi-arabia-robot-citizen-sophia/.
6. He Ren Q. Y. Neural Joke Generation. Department of Electrical Engineering, Stanford University. https://pdfs.semanticscholar.org/e72f/c4aa5076996ab0272a3de291c9ecb88698d4.pdf.
7. Kissinger H. A. How the Enlightenment Ends. www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/.