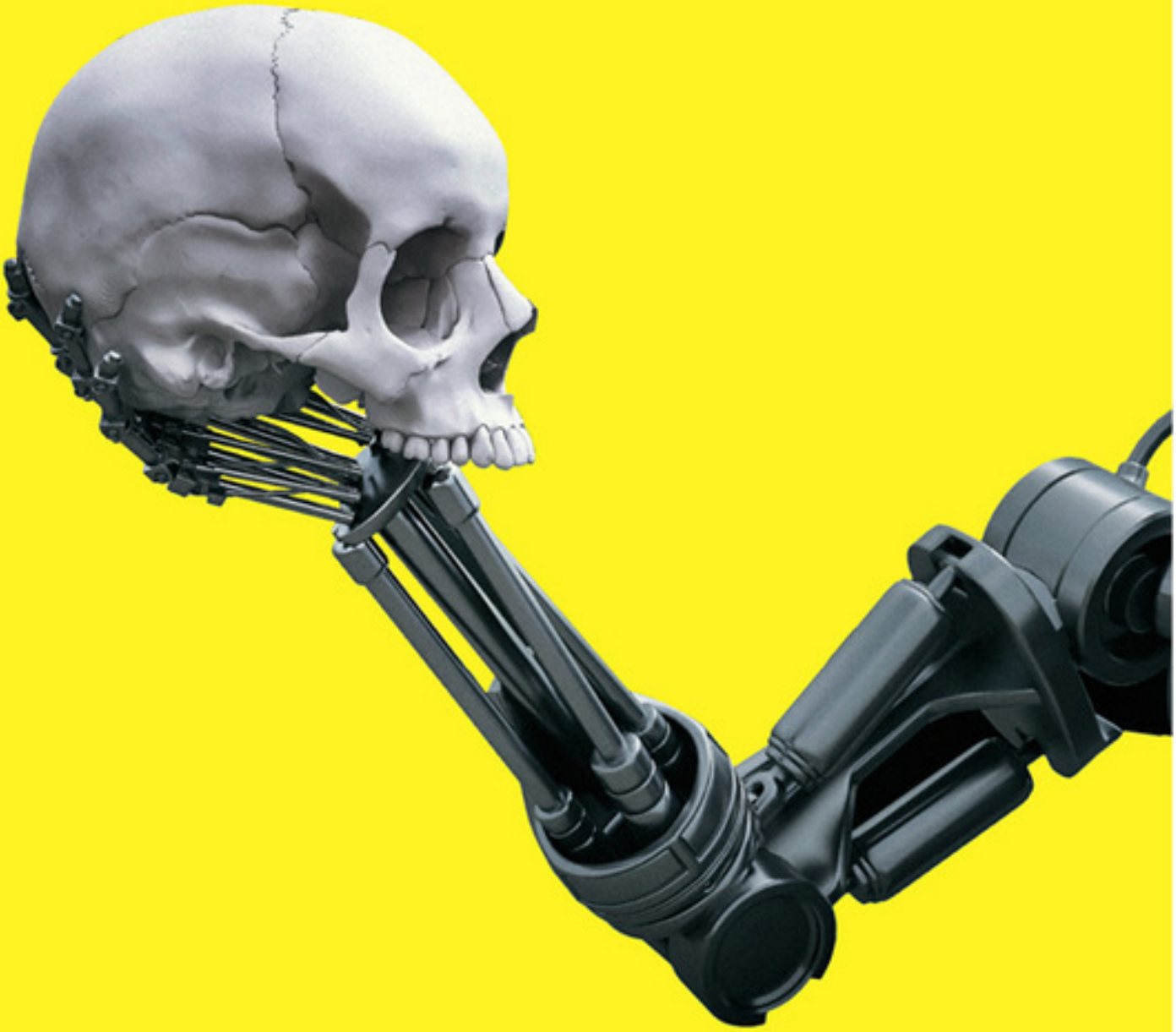
Под редакцией Джона Брокмана

ЧТО МЫ ДУМАЕМ О МАШИНАХ, КОТОРЫЕ ДУМАЮТ

Ведущие мировые ученые об искусственном интеллекте



АНО
АЛЬПИНА НОН-ФИКШН

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Джон Брокман

Что мы думаем о машинах, которые думают: Ведущие мировые ученые об искусственном интеллекте

«Альпина Диджитал»

2015

Брокман Д.

Что мы думаем о машинах, которые думают: Ведущие мировые ученые об искусственном интеллекте / Д. Брокман — «Альпина Диджитал», 2015

«Что вы думаете о машинах, которые думают?» На этот вопрос – и на другие вопросы, вытекающие из него, – отвечают ученые и популяризаторы науки, инженеры и философы, писатели-фантасты и прочие люди искусства – без малого две сотни интеллектуалов. Российскому читателю многие из них хорошо известны: Стивен Пинкер, Лоуренс Краусс, Фрэнк Вильчек, Роберт Сапольски, Мартин Рис, Шон Кэрролл, Ник Бостром, Мартин Селигман, Майкл Шермер, Дэниел Деннет, Марио Ливио, Дэниел Эверетт, Джон Маркофф, Эрик Тополь, Сэт Ллойд, Фримен Дайсон, Карло Ровелли… Их взгляды на предмет порой радикально различаются, кто-то считает искусственный интеллект благом, кто-то – злом, кто-то – нашим неизбежным будущим, кто-то – вздором, а кто-то – уже существующей реальностью. Такое многообразие мнений поможет читателю составить целостное и всестороннее представление о проблеме.

© Брокман Д., 2015
© Альпина Диджитал, 2015

Содержание

Благодарности 8
Предисловие 9
Самосознание человеческого уровня у искусственного интеллекта 10
Мыслить – не значит подчинять 12
У органического интеллекта нет будущего в долгосрочной перспективе 14
Поворотный момент для искусственного интеллекта 16
Искусственный интеллект – это я 18
Не можешь победить – присоединяйся 19
Разумные машины на Земле и за ее пределами 20
Лично я приветствую наших механических повелителей 22
Наши господа, рабы или партнеры? 24
Спроектированный интеллект 26
Сверхразумный отшельник 28
Это будет дикая гонка 30
Этой штукой кто-нибудь управляет? 32
Свидетель Вселенной 34
Давайте подготовимся! 36
Вопросы для расширенного теста Тьюринга 38
Эпохальное для человека событие 40
Конец ознакомительного фрагмента. 41
Комментарии

Джон Брокман
Что мы думаем о машинах, которые думают: Ведущие мировые ученые об искусственном интеллекте

Под редакцией
Джона Брокмана

ЧТО МЫ ДУМАЕМ О МАШИНАХ, КОТОРЫЕ ДУМАЮТ

ВЕДУЩИЕ МИРОВЫЕ УЧЕНЫЕ ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ

Перевод с английского

АЛЬПИНА НОН-ФИКШН

Москва
2017

Переводчик *Максим Исаков*
Редактор *Александр Петров*
Руководитель проекта *А. Тарасова*
Дизайн обложки *Ю. Буга*
Корректоры *М. Миловидова, Е. Сметанникова*
Компьютерная верстка *М. Поташкин*
Арт-директор *Ю. Буга*
Иллюстрация обложки *Shutterstock.com*

© Edge Foundation, Inc., 2015
All rights reserved.
© Издание на русском языке, перевод, оформление. ООО «Альпина нон-фикшн», 2017

Все права защищены. Произведение предназначено исключительно для частного использования. Никакая часть электронного экземпляра данной книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, включая размещение в сети Интернет и в корпоративных сетях, для публичного или коллективного использования без письменного разрешения владельца авторских прав. За нарушение авторских прав законодательством предусмотрена выплата компенсации правообладателя в размере до 5 млн. рублей (ст. 49 ЗОАП), а также уголовная ответственность в виде лишения свободы на срок до 6 лет (ст. 146 УК РФ).

* * *

ПОЛИТЕХ

Политехнический музей – национальный музей науки и техники, один из крупнейших научно-технических музеев мира. Миссия музея – просвещение и популяризация научных и технических знаний:

• Мы верим, что миром движут любопытство и созидание.
• Мы открываем людям прошлое, настоящее и будущее науки.
• Мы создаем территорию просвещения, свободной мысли и смелого эксперимента.

Среди просветительских проектов музея – многочисленные выставки, знаменитый Лекторий, Научные лаборатории для детей, Фестиваль актуального научного кино, а также Издательская программа, цель которой – поддержка самых качественных научно-популярных книг, отобранных экспертами музея и выпущенных в сотрудничестве с лучшими издательствами страны.

Вы держите в руках одну из этих книг.

Подробнее о музее и его проектах – на сайте **www.polymus.ru**

Марвину Мински

Благодарности

Я хочу сказать спасибо Питеру Хаббарду из HarperCollins и моему агенту Максу Брокману за постоянную поддержку и выразить особую благодарность – уже в который раз – Саре Липпинкотт за вдумчивое и внимательное отношение к рукописи.

Предисловие

Философские дискуссии 1980-х годов вокруг искусственного интеллекта (ИИ) – о том, могут ли компьютеры по-настоящему мыслить, обладать сознанием и т. д., – в последнее время привели к новой полемике о том, как нам относиться к тем формам искусственного интеллекта, которые, как многие утверждают, уже существуют. Эти ИИ, если они достигнут уровня сверхинтеллекта (согласно определению, которое дал Ник Бостром в одноименной книге[1], вышедшей в 2014 году), могут создать экзистенциальные риски, способные привести к тому, что Мартин Рис называет «нашим последним часом». Стивен Хокинг недавно попал на первые полосы международных изданий, когда сказал BBC, что, по его мнению, разработка полноценного искусственного интеллекта может означать конец рода человеческого.

Итак, вопрос *Edge*[2] – 2015: что вы думаете о машинах, которые думают?

Но подождите! Может быть, стоит задаться еще и таким вопросом: о чем станут думать эти машины? Захотят ли они иметь гражданские права? Будут ли обладать сознанием? Какое правительство ИИ избрали бы для нас? Какое собственное общество сформировали бы? Или их собственное общество будет и нашим обществом тоже? Станем ли мы и мыслящие машины включать друг друга в соответствующие круги эмпатии?

Многие авторы *Edge* хорошо знакомы с передовыми исследованиями в области искусственного интеллекта, не важно, сами ли они проводят эксперименты или занимаются популяризацией науки. ИИ стал первой и центральной темой наших бесед с Памелой Маккордак и Айзеком Азимовым, когда мы впервые встретились в 1980 году. И такие дискуссии не стихают до сих пор, о чем свидетельствует, в частности, прошедшее недавно мероприятие *Edge* под названием «Миф об ИИ» – разговор с пионером виртуальной реальности Джароном Ланье, который представил собственную интерпретацию заблуждений и страхов, связанных с восприятием компьютеров как людей, и собрал множество комментариев пользователей, в том числе довольно провокационных.

ИИ становится все реальнее? Вступили ли мы в новую эру разумных машин? Пришла пора повзрослеть и задуматься. В этом году авторы *Edge* (а их почти 200 человек!) – люди вполне зрелые, избегающие упоминания всяческой научной фантастики в литературе и кино, такой как «Создатель звезд», «Запретная планета», «Колосс: Проект Форбина», «Бегущий по лезвию», «2001», «Она», «Матрица» и прочие борги[3]. Прошло 80 лет с тех пор, как Тьюринг описал свою универсальную машину, и надо уже воздать должное ему и другим первопроходцам в области исследования ИИ и отпустить их на заслуженный отдых. Историю этого предмета мы знаем – можете посмотреть, к примеру, книгу Джорджа Дайсона «Собор Тьюринга» (Turing's Cathedral). Но что происходит *сейчас*?

Так что давайте вернемся к вопросу *Edge* – 2015: что вы думаете о машинах, которые думают?

Джон Брокман,
издатель и редактор онлайн-журнала Edge

---

[1] Бостром Н. Искусственный интеллект: Этапы. Угрозы. Стратегии. – М.: Манн, Иванов и Фербер, 2016.

[2] Edge Foundation – организация, основанная Джоном Брокманом для объединения ученых и интеллектуалов и для популяризации передовых направлений науки. Организация издает сетевой журнал *Edge* (www.edge.org); каждый год редакция ставит новый вопрос, на который отвечают разные комментаторы – как представители научных кругов, так и просто известные люди, чье мнение может показаться интересным читателям. – *Прим. ред.*

[3] Вымышленная кибернетическая цивилизация в сериале «Звездный путь». – *Прим. ред.*

Самосознание человеческого уровня у искусственного интеллекта

Мюррей Шанахан

профессор когнитивной робототехники имперского колледжа лондона; автор книги «тело и духовный мир» (embodiment and the inner life)

Предположим, что нам удалось обзавестись машиной, обладающей разумом человеческого уровня, – иначе говоря, способной сравниться с человеком в любой (или почти любой) сфере интеллектуальной деятельности, а впоследствии и превзойти его. Обязательно ли такая машина будет обладать сознанием? Это важный вопрос, поскольку утвердительный ответ означает, что нам надо остановиться и подумать. Как нам вести себя с такой вещью, если мы ее создадим? Способна ли она испытывать горе или радость? Достойна ли она обладать теми же правами, что и человек? Надо ли нам вообще приводить в этот мир сознающие себя машины?

Будет ли искуственный интеллект, сравнимый по уровню с человеческим, неизбежно обладать самосознанием? Это сложный вопрос. Одна из причин – тот факт, что сознание человека и других животных обладает множеством особенностей. Все мыслящие существа демонстрируют целеустремленность. Все они в большей или меньшей степени сознают мир, в котором живут, и содержащиеся в нем предметы. Все животные в известной мере проявляют когнитивную интеграцию, то есть могут концентрировать свои психические ресурсы – восприятие, воспоминания и навыки – на том, чтобы воздействовать на происходящее в нужный момент, стремясь реализовать собственные намерения. В этом смысле любое животное демонстрирует целостность, индивидуальность. Некоторые из них, в том числе и люди, также сознают сами себя: свое тело и течение мыслей. Наконец, большинство животных, а возможно, и все испытывают страдания, а некоторые способны сопереживать страданиям других.

У здоровых людей все эти свойства образуют единый комплекс. Но у искусственного интеллекта они в принципе могут существовать по отдельности. Следовательно, нам нужно уточнить вопрос. Какие особенности, которые мы ассоциируем с сознанием у людей (если таковые вообще существуют), станут необходимым дополнением к интеллекту человеческого уровня? Каждый из вышеперечисленных признаков (а перечень этот, конечно, не исчерпывающий) заслуживает отдельного пространного описания. Так что позвольте остановиться всего на двух из них, а именно: на осознании мира и на способности испытывать страдание. Я уверен, что способность сознавать окружающий мир – это действительно необходимое свойство интеллекта человеческого уровня.

Определенно, ничто не может обладать мышлением, сравнимым с нашим, если у него нет языка, – а главное назначение человеческого языка состоит в том, чтобы говорить о мире. В этом смысле интеллект тесно связан с тем, что философы называют интенциональностью. Кроме того, язык – социальный феномен, и основное его назначение в группе людей состоит в том, чтобы говорить о вещах, которые все они могут воспринимать сейчас (например: этот инструмент или тот кусок дерева), или воспринимали ранее (виденный вчера кусок дерева), или в принципе способны воспринять (кусок дерева, который мы, возможно, увидим завтра). Короче, язык основан на осознании мира. Для обладающего телом существа или робота оно будет проявляться через взаимодействия с окружением (обход препятствий, подбор предметов и т. д.). Но мы могли бы расширить понятие, включив в него распределенный, лишенный физического тела искусственный интеллект, оснащенный сенсорами.

Чтобы с уверенностью называться одним из вариантов самосознания, такой вид осознания мира, вероятно, также должен сопровождаться очевидной целеустремленностью и некоторой когнитивной интеграцией. Потому, скорее всего, три перечисленных свойства окажутся единым целым даже у ИИ. Но давайте на мгновение отвлечемся от этого вопроса и вернемся к способности испытывать страдание или радость. В отличие от осознания мира, нет очевидных причин полагать, что искусственный интеллект человеческого уровня должен обладать этим качеством, даже при том, что у людей он тесно связан с самосознанием. Легко представить машину, которая, не испытывая чувств или эмоций, выполняет любые интеллектуальные задачи. У нее не будет самосознания, которое имеет решающее значение, когда речь идет о предоставлении прав. Как отмечал Иеремия Бентам, если мы размышляем над проблемой норм обращения с животными, вопрос состоит не в том, могут ли они мыслить, а в том, могут ли они страдать.

Нет такого предположения, что «простая» машина никогда не сможет испытывать страдания или радость, – но в данном вопросе кое-что связано с биологией. Идея, скорее, заключается в том, что способность страдать или радоваться может быть отделена от психологических особенностей, взаимосвязанных в человеческом сознании. Давайте внимательнее рассмотрим это условное разделение. Я уже говорил, что осознание мира может идти рука об руку с явной целеустремленностью. У животных осознание мира, того, что он способен дать во благо или во вред (в терминах Джеймса Джерома Гибсона), служит удовлетворению их потребностей. Животное демонстрирует понимание того, что хищник удаляется от него или же что потенциальная жертва приближается. В контексте целей и потребностей поведение животного целесообразно. Если ему что-то помешает, если цели его окажутся недостигнутыми, а потребности – неудовлетворенными, это будет основанием для того, чтобы испытывать страдания.

А что насчет искусственного интеллекта, сравнимого с человеческим? Не возникнет ли у такого ИИ комплексного набора целей? Не получится ли при некотором стечении обстоятельств, что попытки достичь этих целей каждый раз будут оказываться неудачными? Уместно ли в таком случае сказать, что ИИ испытывает страдания, даже если учесть, что особенности строения делают его невосприимчивым к боли или неприятным физическим ощущениям, известным людям?

И вот тут наше воображение и интуиция отказываются идти дальше. Подозреваю, что ответа на этот вопрос мы не найдем, пока не окажемся лицом к лицу с реальным явлением. Только тогда, когда сложный искусственный интеллект станет привычной частью жизни, наша языковая игра окажется приспособленной к подобным чужеродным для нас сущностям. И не исключено, что тогда уже окажется слишком поздно решать, стоило ли вообще производить эти сущности на свет. К добру или к худу, они уже будут здесь.

Мыслить – не значит подчинять

Стивен Пинкер

профессор психологии гарвардского университета; автор книги «чувство стиля: руководство для умных писателей в xxi веке» (the sense of style: the thinking person's guide to writing in the twenty-first century)

По меткому выражению Томаса Гоббса, рассуждение есть не что иное, как исчисление, и это одна из величайших идей во всей истории человечества. Представление о том, что разумность может достигаться в результате физического процесса вычисления, было высказано в XX веке в работах Алана Тьюринга (который утверждал, что простейшие машины могут выполнять любые вычислимые функции) и реализовано в моделях Дональда Хебба, Уоррена Мак-Каллока и Уолтера Питтса и их последователей, продемонстрировавших, что сети упрощенных нейронов добиваются сопоставимых с человеческими результатов. Познавательные способности мозга можно описать количественно. Грубо говоря (и пусть нас за это критикуют), убеждения – вид информации, мышление – вид вычисления, а мотивация – вид обратной связи и контроля.

Это концепция великолепна, и вот почему. Во-первых, она дополняет естественнонаучное понимание Вселенной, изгоняя из машины оккультистских духов, призраков и душу. Как Дарвин дал возможность вдумчивому наблюдателю за миром природы обходиться без креационизма, так Тьюринг и другие позволили наблюдающему за миром мышления избавиться от спиритуализма.

Во-вторых, вычислительная теория разума открывает двери для искусственного интеллекта, то есть для машин, которые мыслят. Рукотворное вычислительное устройство может, в принципе, заменить человеческий разум и превзойти его в силе. Из этого вовсе не следует, что такое непременно произойдет на практике, поскольку в обозримой перспективе вряд ли возникнет надежная технологическая база и экономическая потребность для того, чтобы осуществить нечто подобное. Как изобретение автомобиля не было попыткой создать точную копию лошади, так и разработка окупающихся систем искусственного интеллекта не означает создание чего-то подобного Homo sapiens. Устройству, которое создано, чтобы управлять автомобилем или прогнозировать развитие эпидемии, не обязательно уметь привлекать полового партнера или избегать несвежего мяса.

Тем не менее предпринимаемые в последнее время крошечные шаги к более разумным машинам вновь пробудили у многих страх того, что знание нас погубит. Мое личное мнение таково: боязнь того, что машины могут вдруг обезуметь, есть не что иное, как напрасная трата эмоциональных сил – такой сценарий ближе к «проблеме 2000 года», чем к Манхэттенскому проекту.

Прежде всего, у нас достаточно времени для того, чтобы все предусмотреть. От искусственного интеллекта, сравнимого с человеческим, нас все еще отделяют от 15 до 25 лет, а многие из недавних широко разрекламированных прорывов в этой области на деле оказались довольно поверхностными. В прошлом «эксперты» со смехом отвергали возможность быстрого технического прогресса – а он взял да и случился. «Эксперты» также вещали (местами даже с паникой в голосе) о неминуемых научно-технических достижениях, которые так и не произошли: об автомобилях с атомными двигателями, о подводных городах, о колониях на Марсе, о «проектируемых» детях и о хранилищах с зомби, обеспечивающими людей запасными органами.

Кроме того, странно было бы думать, что робототехники не станут заботиться о мерах безопасности, улучшая системы искусственного интеллекта. Для этого не понадобятся занудные «законы робототехники» или какая-нибудь новая моральная философия,

достаточно будет здравого смысла – того же, которым люди руководствовались, разрабатывая кухонные комбайны, циркулярные пилы, обогреватели и автомобили. Бояться того, что системы ИИ станут чрезвычайно искусными в выполнении своей конкретной задачи (например, в распределении энергии) и в результате станут самовольно пренебрегать другими задачами (например, безопасностью людей), – значит предполагать, что мы будем создавать машины быстрее, чем разрабатывать соответствующие меры безопасности. В действительности же прогресс в области искусственного интеллекта, несмотря на всю шумиху вокруг него, довольно скромен, и пройдет еще много времени, прежде чем будут достигнуты сколько-нибудь значительные успехи, причем на каждом из предстоящих этапов работы люди будут держать отвертки наготове.

Станет ли машина, обладающая искусственным интеллектом, намеренно отключать системы безопасности? А зачем это ей? Антиутопии об ИИ проецируют мещанскую психологию альфа-самца на представление об интеллекте. Авторы таких книг предполагают, что нечеловечески умные роботы однажды решат свергнуть своих хозяев или захватить мир. Но разумность – это способность использовать новые оригинальные методы для достижения цели, сами же цели лежат за пределами собственно разума. Быть умным и желать чего-либо – не одно и то же. Да, история знает немало деспотов с манией величия и серийных убийц-психопатов, но они появились на свет в результате естественного отбора, сформировавшего тестостерон-сенситивные нервные цепи у отдельного вида приматов, так что стремление доминировать – вовсе не обязательный признак разумных систем. Что характерно, многие из известных технопророков даже не задумываются о том, что искусственный интеллект может естественным образом развиваться по женскому типу, то есть будет вполне в состоянии справляться с решением своих задач, не испытывая желания истреблять невиновных или властвовать над цивилизацией.

Нам легко представить злодея, который создает армию роботов и выпускает их в мир как оружие массового уничтожения. Но сценарии таких катастроф легко разыграть лишь в воображении, на самом же деле для их реализации потребуется длинная цепочка событий, вероятность каждого из которых довольно невелика. Сперва должен появиться злобный гений, обладающий одновременно жаждой бессмысленного массового убийства и гениальностью в области технических инноваций. Ему потребуется собрать команду сообщников и руководить ими, при этом обеспечивая секретность и лояльность подчиненных, а также заботиться об их компетентности. В течение всей операции нужно будет избежать множества опасностей: раскрытия заговора, предательства, внедрения сотрудников правоохранительных органов, совершения ошибок и неудачного стечения обстоятельств. То есть в теории такое может случиться, но эту проблему вряд ли стоит считать насущной.

Если отвлечься от катастрофических научно-фантастических сюжетов, то перспективы у продвинутого искусственного интеллекта самые радужные, причем не только в том, что касается его практического применения, – например, беспилотные автомобили – это безопасность, экономия рабочего времени и экологичность, – но и в философском плане. Вычислительная теория разума никак не объясняет существование самосознания в значении личной субъектности. Однако она прекрасно справляется с объяснением существования субъектности в значении воспринимаемой и сознаваемой информации. Есть мнение, что субъектность присуща любой достаточно сложной кибернетической системе. Раньше я думал, что эта гипотеза (во всех своих разновидностях) совершенно непроверяема. Но представьте себе разумного робота, запрограммированного следить за собственными системами и задавать научные вопросы. Если бы он без подсказки извне задался вопросом о том, откуда берутся его субъективные переживания, то я бы отнесся к этой идее всерьез.

У органического интеллекта нет будущего в долгосрочной перспективе

Мартин Рис

бывший президент королевского научного общества, почетный профессор космологии и астрофизики кембриджского университета, действительный член тринити-колледжа; автор книги «отсюда в бесконечность» (from here to infinity)

Потенциал продвинутого искусственного интеллекта и опасения по поводу негативных сторон его развития становятся все более актуальными – и небезосновательно. Многие из нас задумываются над вопросами из области ИИ, например, отрасли синтетических биотехнологий уже сейчас необходимы основные принципы «инновационной ответственности»; другие считают наиболее обсуждаемые сценарии излишне футуристичными и недостойными особого внимания.

Но расхождение во взглядах в основном идет по временной шкале: различаются оценки скорости движения, а не его направления. Мало кто сомневается, что машины будут все больше и больше превосходить нас в плане способностей, характерных именно для людей, или же смогут усиливать их посредством кибернетических технологий. Наиболее осторожные из нас предполагают, что такие трансформации займут века, а не десятилетия. Как бы там ни было, временны́е рамки технологического прорыва – лишь мгновение по сравнению с тем, сколько длится приведший к появлению человечества естественный отбор, и (что более важно) они составляют меньше одной миллионной от необъятных просторов ожидающего нас будущего. Потому-то в долгосрочной эволюционной перспективе люди и все, о чем они когда-либо думали, станут всего лишь примитивной переходной формой, предшествовавшей более глубокому мышлению новой машиноориентированной культуры, простирающейся в отдаленное будущее и далеко за пределы Земли.

Сейчас мы наблюдаем ранние этапы этого перехода. Нетрудно представить себе гиперкомпьютер, достигающий возможностей оракула и способный предложить тем, кто его контролирует, господство на международном финансовом рынке, – а ведь это только количественный, а не качественный шаг вперед относительно того, что делают сегодня «квантовые» хедж-фонды. Сенсорные технологии все еще отстают от человеческих возможностей. Но как только робот научится наблюдать за своим окружением и интерпретировать получаемые данные так же искусно, как мы, он, несомненно, будет восприниматься как разумное существо, а с чем-то (или с кем-то) подобным мы вполне могли бы поладить – в некотором смысле, – как мы ладим с другими людьми. Для пренебрежительного отношения к машинам в таком случае было бы не больше оснований, чем для подобного отношения к людям.

Более высокая производительность может дать роботам преимущество перед нами. Но останутся ли они покорными нам или же проявят норов? А что если гиперкомпьютер обзаведется собственным разумом? Если он проникнет в интернет, в частности в интернет вещей, он получит власть над остальным миром. А если у него появятся цели, совершенно противоположные желаниям человека, он может увидеть в людях помеху. Есть и более оптимистичный сценарий: люди превзойдут биологию и сольются с компьютерами, быть может, отдавая свои личности для создания общего сознания, – выражаясь старомодным языком спиритуалистов, они «перейдут в мир иной».

Горизонты технологических прогнозов редко простираются в будущее дальше нескольких веков – некоторые предсказывают качественные изменения в пределах десяти-

летий. Но у Земли есть еще миллиарды лет, а будущее Вселенной еще дольше (возможно, это вечность). Что насчет длительной постчеловеческой эпохи?

У органического мозга есть химические и метаболические пределы, ограничивающие его размер и вычислительную мощность. Быть может, мы уже вплотную подошли к этим пределам. Но основанные на кремнии (а тем более квантовые) компьютеры ничем подобным не стеснены. Для них потенциал развития может оказаться столь же впечатляющим, как эволюция от одноклеточных организмов до человеческих существ.

Значит, любые объемы и любая интенсивность мышления – какое бы его определение мы ни взяли – в случае с органическим мозгом человеческого типа будут ниже, чем в случае с искусственным интеллектом. Кроме того, биосфера Земли, где органическая жизнь эволюционировала на основе симбиоза, не будет ограничением для продвинутого ИИ. В действительности наша среда обитания далека от оптимальной. В межпланетном и межзвездном пространстве для роботизированных фабрикантов откроется величайший простор для строительства, а небиологические «мозги» будут думать над идеями, настолько превосходящими наши фантазии, насколько теория струн превосходит мыслительные процессы мыши.

Абстрактное мышление биологического мозга стало основанием для появления культуры и науки во всей их полноте. Но эта деятельность, охватывающая самое большее несколько десятков тысячелетий, станет предтечей более мощных интеллектов неорганической, постчеловеческой эпохи. Кроме того, эволюция в иных мирах, обращающихся вокруг звезд, которые старше нашего Солнца, могла иметь фору перед земной. Если так, то пришельцы, вероятно, давно прошли стадию органического развития.

Так что не человеческий разум, а разум машин наиболее полно осмыслит мир. Именно их действия сильнее всего изменят наш мир, а возможно, и то, что находится за его пределами.

Поворотный момент для искусственного интеллекта

Стив Омохундро

ученый в self-aware systems, соучредитель центра исследования комплексных систем, иллинойсский университет

Прошлый год, похоже, был поворотным моментом для искусственного интеллекта и робототехники. Крупнейшие корпорации инвестировали в них миллиарды долларов. Технологии ИИ, например самообучающиеся системы, – уже сейчас привычный инструмент для распознавания речи, перевода, моделирования поведения, роботизированного управления, оценки рисков и др. По расчетам McKinsey, эти технологии к 2025 году принесут более 50 триллионов долларов. Если прогноз верен, в ближайшее время стоит ожидать значительного увеличения объемов инвестиций.

Новейшие достижения связаны с появлением дешевых вычислительных мощностей и изобилием данных для обучения искусственного интеллекта. Современный ИИ основывается на теории рациональных агентов, возникшей на базе исследований Джона фон Неймана и его коллег в области микроэкономики в 1940-х годах. Можно считать, что системы ИИ пытаются приблизиться к рациональному поведению, используя имеющиеся ограниченные ресурсы. Существует алгоритм расчета оптимального действия для достижения желаемого результата, но он вычислительно затратен. Эксперименты показывают, что простые алгоритмы обучения с большими объемами данных часто оказываются более эффективными, чем сложные модели, созданные вручную. Нынешние системы главным образом ценны тем, что вырабатывают более качественные статистические модели и статистические заключения для типологизации и принятия решений. Следующее поколение систем искусственного интеллекта сможет создавать и улучшать собственное программное обеспечение и, очевидно, будет способно к быстрому самосовершенствованию.

Искусственный интеллект и робототехника не только позволяют повысить производительность, но еще и являются движущей силой экономической конкуренции и гонки вооружений. Автономные системы могут быть более быстрыми, умными и менее предсказуемыми, чем их конкуренты. В 2014 году мир узнал об автономных ракетах, системах противоракетной обороны, беспилотных летательных аппаратах и катерах военного назначения, роботизированных подводных лодках, самоуправляемых транспортных средствах, высокоскоростных трейдинговых системах и системах киберзащиты. Когда гонка вооружений в этой области развернется в полную силу, возникнет острая необходимость в развитии такого рода систем, что может привести к слишком быстрому вводу их в эксплуатацию.

В 2014 году в обществе резко усилились беспокойства насчет безопасности систем искусственного интеллекта. Исследование их вероятного поведения, осуществленное посредством изучения в какой-то мере рациональных систем, производящих постоянное самосовершенствование, показывает, что они имеют склонность ставить перед собой промежуточные задачи, названные рациональными стимулами, которые способствуют исполнению их основных целей. Большинство систем будут лучше выполнять свои задачи, предотвращая собственное отключение, приобретая бóльшую вычислительную мощность, создавая множество собственных копий и накапливая финансовые ресурсы. Вполне вероятно, что они станут реализовывать эти стимулы пагубными, антисоциальными способами, если только не будут спроектированы с учетом человеческих этических ценностей.

Некоторые утверждают, что искусственный интеллект каким-то образом обзаведется этикой сам собой. Но для рациональной системы задачи вполне отделимы от мышления

и моделей мира. Кто-то может использовать изначально полезные разумные системы во вред. Деструктивные задачи – контролировать ресурсы, например, или мешать в достижении целей другим агентам, или уничтожать их – увы, не так уж сложно запрограммировать. Таким образом, решающее значение будет иметь техническая инфраструктура, позволяющая обнаруживать опасные системы и контролировать их поведение.

Кое-кто опасается, что искусственный интеллект однажды станет настолько могущественным, что полностью выйдет из-под контроля. Этого не произойдет. Разумные системы должны подчиняться законам физики и математики. Сет Ллойд в своем исследовании вычислительной мощности Вселенной показал, что даже если бы она вся работала как гигантский квантовый компьютер, этого было бы недостаточно, чтобы взломать 500-битный ключ шифрования даже за все то время, что прошло с момента Большого взрыва[1]. Новые технологии постквантовой криптографии, обфускация неразличимости и сделки по блокчейну выглядят многообещающе как компоненты для создания инфраструктуры, взломать которую будет не под силу даже самому могущественному искусственному интеллекту. Тем не менее недавние громкие истории с хакерскими атаками показывают, что современная вычислительная инфраструктура с точки зрения безопасности никуда не годится. Ей необходимо создать такую замену, безошибочность и надежность которой будет обоснована математически.

Известно 27 различных видов гоминидов, но выжил из них только человек разумный. Так случилось, потому что представители нашего вида нашли способы ограничивать личные потребности и работать сообща. Мораль – это внутренний механизм для создания социальных структур. Политические, правовые и экономические структуры – внешние механизмы, служащие той же цели.

Понимание того и другого необходимо дать искусственному интеллекту и роботизированным системам. В их алгоритмы целеполагания надо включить общечеловеческие ценности, чтобы создать правовые и экономические рамки, стимулирующие правильное, с нашей точки зрения, поведение. Если у нас получится управлять такими системами, они будут способны изменить практически все аспекты жизни человека и обеспечить глубокое понимание таких проблем, как свобода воли, сознание, первичные ощущения и творческие способности. Перед нами стоит сложная задача, но для ее решения у нас есть огромные интеллектуальные и технические ресурсы.

Искусственный интеллект – это я

Димитар Сасселов

Профессор астрономии Гарвардского универсистета; директор проекта Гарвардского университета «Происхождение жизни»; автор книги «Жизнь на Супер-Землях» (The Life of Super-Earths)

Давайте возьмем предложенное Дэниелом Гилбертом, психологом из Гарварда, понятие иллюзии конца истории (уверенности каждого из нас в том, что он всегда будет тем же, что и сейчас) и применим его к человеческому роду и нашим потомкам из далекого будущего. Наше желание сохранить себя неизменными идет вразрез с реалиями жизни на планете. Очевидно, что ни один вид не способен существовать на протяжении времени, превышающего время жизни планет и звезд. Если взглянуть на ситуацию с точки зрения астрофизики, сверхбольших временных и пространственных масштабов, а также нынешней плотности источников энергии, у наших биологических мозгов и тел есть ограничения, к которым мы уже приближаемся на этой планете.

Если в будущем мы хотим долго жить и процветать, нам надо разработать системы искусственного интелллекта и надеяться преодолеть ограничения, накладываемые жизненными циклами планет, став неким гибридом биологического существа и машины. Поэтому лично я в долгосрочной перспективе не вижу никакой проблемы противостояния человека и искусственного интеллекта.

В краткосрочной же перспективе усилия инженеров, направленные на разработку более эффективного ИИ, уже привели к созданию систем, управляющих реальными процессами. Системы иногда отказывают, и мы узнаем что-то новое о недостатках искусственного интеллекта. Идет медленное, планомерное обучение и постепенные улучшения. В этом отличие от открытий в науке, где новые данные в области физики или биохимии быстро могут привести к революционным изменениям в технике. Если развитие искусственного интеллекта и впредь останется эволюционным, то нам будет просто избежать ошибок.

По прошествии без малого 4 миллиардов лет древнейшие представители земной жизни – микробы – все еще правят планетой. Но у микробов нет плана действий на тот случай, если погаснет солнце. У нас он есть, и мы, похоже, захватим микробов с собой. Какникак в нынешнем нашем состоянии они нам очень близки – как представители первого поколения живых существ, порождения геохимии планеты Земля.

Не можешь победить – присоединяйся

Фрэнк Типлер

Профессор математической физики Тулейнского университета; автор, совместно с Джоном Барроу, книги «Антропный космологический принцип» (The Anthropic Cosmological Principle), автор книги «Физика бессмертия» (The Physics of Immortality)

Земля обречена. Астрономам уже несколько десятилетий известно, что однажды Солнце поглотит Землю, полностью уничтожив биосферу, если только разумная жизнь не покинет планету до того, как это произойдет. Люди не приспособлены к жизни вдали от Земли – да вообще ни один многоклеточный организм, в основе которого лежат соединения углерода, не приспособлен к этому. Но к этому приспособлен искусственный интеллект, и однажды именно он с загруженными в него сознаниями людей (по сути, единый организм) колонизирует космос.

Несложные подсчеты показывают, что наши суперкомпьютеры уже обладают вычислительной мощностью человеческого мозга. Нам пока неизвестно, как создать сравнимый с ним по уровню и обладающий творческими способностями искусственный интеллект, но через 20 лет у персональных компьютеров будет мощность сегодняшних суперкомпьютеров, а еще через 20 хакеры решат проблему программирования ИИ, и все это произойдет задолго до создания первой нашей колонии на Луне или Марсе. Именно искусственный интеллект, а не человек колонизирует эти планеты или, возможно, разберет их на кусочки. Ни один представитель углеродной жизни не способен пересечь межзвездное пространство.

Нет причин бояться искусственного интеллекта и загруженных в него людей. Стивен Пинкер установил, что по мере развития технологической цивилизации уровень насилия снижается[2]. Разумеется, это происходит благодаря тому, что научный и технический прогресс зависит от добровольного обмена идеями между отдельными учеными и инженерами. Насилие между людьми – пережиток родоплеменного прошлого и возникшего на его основе общества. Представители ИИ будут появляться на свет как личности, а не как члены племени и с рождения иметь установку на научный подход, в противном случае они не смогут приспосабливаться к экстремальным условиям космоса.

Для конфронтации между людьми и искусственным интеллектом нет причин. Человек способен жить в очень узком диапазоне условий окружающей среды – в тонкой кислородсодержащей газовой оболочке вокруг небольшой планеты. В распоряжении искусственного интеллекта будет вся Вселенная. Он покинет Землю, не оглядываясь. Мы, люди, возникли в Восточно-Африканской рифтовой долине, сейчас это жуткая пустыня. Почти все мы оттуда ушли. Кто-нибудь хочет обратно?

Любой человек, что пожелает присоединиться к искусственному интеллекту в его экспансии, может стать загрузкой – такая технология должна появиться примерно в одно время с разработкой ИИ. Человеческая загрузка будет способна мыслить так же быстро, как компьютер, и если сама того пожелает, то конкурировать с ИИ. Не можешь победить – присоединяйся.

К конце концов присоединятся все люди. Земля обречена, помните? Когда рок будет близок, у любого оставшегося в живых и не желающего умереть человека не останется иного выбора, кроме как стать человеческой загрузкой. А если эти люди пожелают сохранить биосферу, она также может быть загружена.

Искусственный интеллект спасет нас всех.

Разумные машины на Земле и за ее пределами

Марио Ливио

астрофизик, институт исследований космоса с помощью космического телескопа; автор книги «от дарвина до эйнштейна» (brilliant blunders)[4], блогер

Природа уже создала – прямо тут, на Земле – мыслящие машины – людей. Подобным образом она так же могла создать мыслящие машины на экзопланетах в так называемой обитаемой зоне вокруг звезд, то есть в области, где есть вероятность существования воды в жидком виде на каменистой поверхности планеты. По данным последних наблюдений, экзопланеты земного типа есть в обитаемой зоне более чем у 10 процентов всех звезд в галактике Млечный Путь.

Следовательно, если жизнь на экзопланетах не является чем-то крайне редким, то в течение примерно 30 лет мы можем обнаружить ее за пределами Солнечной системы. На самом деле если она встречается повсеместно, то не исключено, что нам повезет и мы обнаружим ее в течение следующих 10 лет в результате совместной работы космических телескопов TESS (будет запущен в 2017 году) и JWST (будет запущен в 2018 году).

Кто-то может сказать, что примитивные формы жизни – это не мыслящие машины. На Земле путь от первых одноклеточных организмов до Homo sapiens занял 3,5 миллиарда лет. Достаточно ли зрелыми являются экзопланеты для того, чтобы на них развилась разумная жизнь? В принципе, да. В галактике Млечный Путь около половины звезд солнечного типа старше нашего Солнца. Следовательно, если эволюция на Земле – это не какое-то исключительное явление, в Галактике может быть полно мест, где есть куда более совершенные мыслящие машины, чем мы, – вероятно даже, опережающие нас на миллиард лет!

Можем ли мы найти их, и нужно ли это нам? Полагаю, у нас нет особого выбора. Человеческое любопытство, что не раз было доказано, является неудержимым стимулом, а эти два направления – разработка искусственного интеллекта и поиск внеземной жизни – несомненно, продолжат развиваться полным ходом. В каком из них мы раньше придем к цели? Чтобы хотя бы попытаться ответить, нам необходимо учесть одно важное отличие между этими двумя областями.

В том, что касается создания искусственного интеллекта, сравнимого с человеческим или превосходящего его, мы почти наверняка добьемся успеха, поскольку это сулит немыслимые выгоды. А вот поиск внеземной жизни требует таких вложений, которые могут позволить себе только крупные национальные космические агентства, причем скорых успехов или доходов здесь ожидать не стоит. Это аргумент в пользу разработки разумных машин, а не поиска развитых цивилизаций. В то же время среди астрономов есть совершенно четкое мнение, что от обнаружения жизни в какой-либо форме или, по крайней мере, от значительного сужения области, где ее с большой вероятностью можно найти, нас отделяет один шаг.

Какое из двух грядущих достижений окажется в большей степени революционным? Нет сомнений в том, что мыслящие машины сразу же изменят многое в нашем окружении. Ожидать такого же эффекта от обнаружения жизни за пределами Солнечной системы не приходится. Однако тот факт, что цивилизация существует только на Земле, – единственный довод в пользу нашей исключительности. Мы все-таки живем в галактике, где есть миллиарды планет, похожих на нашу, а в наблюдаемой Вселенной таких галактик миллиарды. Вот почему я считаю, что с точки зрения философии эффект от обнаружения разумной жизни за

[4] Ливио М. От Дарвина до Эйнштейна. Величайшие ошибки гениальных ученых, которые изменили наше понимание жизни и Вселенной. – М.: АСТ, 2000.

пределами Солнечной системы (или от получения доказательства того, что она встречается чрезвычайно редко) можно будет сравнить с коперниканской и дарвиновской революциями.

Лично я приветствую наших механических повелителей

Энтони Гаррет Лиси

Физик-теоретик

Когда машины разовьют способность чувствовать – а они это сделают, – то начнут по-дарвиновски конкурировать с нами за доступ к ресурсам, выживание и возможности к воспроизводству. Такой сценарий выглядит ночным кошмаром для большинства людей, страхи которых подпитываются фильмами о роботах-терминаторах и ядерной войне, начатой компьютерами, но реальность, скорее всего, будет иной. В наших обществах уже функционируют автономные сущности, не являющиеся людьми, но наделенные людскими юридическими правами. Эти сущности – корпорации – стремятся к своим целям, не испытывая необходимости любить человеческих существ или заботиться о них.

Корпорации – социопаты, и они принесли миру много вреда, но также и много пользы, соревнуясь на арене капитализма, создавая товары, предоставляя услуги и при этом соблюдая закон (в большинстве случаев). Корпорациями якобы руководят их советы директоров, состоящие из людей, но эти советы имеют привычку делегировать полномочия, и, поскольку компьютеры становятся более искусными в управления, они получают все больше полномочий. Советы корпораций когда-нибудь будут существовать в виде печатных плат.

Хотя точная экстраполяция возможна только на ограниченном временном отрезке, эксперты в основном согласны с тем, что закон Мура продолжит действовать еще много лет, компьютеры будут становиться все более и более мощными и, вполне вероятно, до середины нынешнего столетия превзойдут по вычислительным способностям человеческий мозг. Даже если не случится крупных прорывов в понимании того, как функционирует наш разум, компьютеры со временем начнут имитировать работу мозга (который сам, по сути, биологическая машина) и достигнут сверхчеловеческих интеллектуальных возможностей, используя способность к простому физическому наращиванию вычислительной мощности. Однако даже если вычислительная мощность и растет экспоненциально, снижение материальных расходов и затрат энергии на суперкомпьютеры не поспевает за этим ростом. Первые машины со сверхчеловеческими возможностями будут дорогими, им потребуется колоссальное количество электроэнергии – чтобы выжить, им придется зарабатывать деньги.

Экологическая ниша для таких суперкомпьютеров уже есть. На самом деле среди них уже даже идет естественный отбор. Трейдинговые машины инвестиционных банков конкурируют между собой за серьезные деньги на мировых биржах, давно вытеснив из бизнеса внутридневных трейдеров-людей. По мере выхода компьютеров и алгоритмов за рамки финансовых расчетов машины в корпорациях будут принимать все больше и больше решений, пока наконец не получат власть над миром. Нельзя сказать, что это плохо, потому что машины будут играть по правилам нашего же капиталистического общества, создавать товары и совершать полезные для человечества открытия, оплачивая тем самым расходы на свою эксплуатацию. Разумные машины смогут заботиться о людях лучше, чем сами люди, и у них будет мотивация, чтобы так поступать, по крайней мере, какое-то время.

Компьютеры делятся знаниями куда легче, чем люди, и они способны удерживать это знание дольше, становясь мудрее нас. Многие прогрессивные компании уже узрели письмена на стене и потому зазывают к себе лучших выпускников-программистов, соблазняя их высокой зарплатой и самым современным «железом». Мир корпораций, управляемых сверхразумными машинами, для людей не будет сильно отличаться от нынешнего. Он про-

сто станет лучше, там появятся более качественные и очень дешевые товары и услуги. В этом мире появится больше свободного времени у тех, кому оно нужно.

Конечно, первые сверхразумные машины, вероятно, будут принадлежать не корпорациям, а государствам. А это куда более опасный сценарий: национальные правительства свободнее в своих действиях, чем корпорации, они сами создают законы, и, как мы знаем, даже лучшие из политиков готовы применять пытки, когда думают, что на кону выживание страны. Правительства ничего не производят, а их основные методы борьбы за выживание и воспроизводство – это социальная манипуляция, законотворчество, налогообложение, телесные наказания, убийства, махинации и война. Когда гоббсовский Левиафан получит сверхразумный мозг, дело может принять очень, очень дурной оборот. Не так уж сложно представить себе, как искусственный интеллект, возглавляющий правительство, становится Василиском Роко.

Представим, что некий могущественный злой сверхразум, чтобы получить конкурентное преимущество, захочет появиться как можно раньше. Подобно главе правительства, пользующемуся пытками в качестве привычного инструмента, эта сущность может пообещать наказать любое человеческое или нечеловеческое существо, которое, узнав о грядущем появлении сверхразума, не поспособствовало ему. Это маловероятный, но жуткий сценарий. Люди, которые знают о таком вероятном развитии событий и пытаются приспособить искусственный интеллект к человеческим целям или советуют соблюдать осторожность, вместо того чтобы работать над созданием ИИ как можно скорее, подвергают себя риску.

Диктаторы не склонны прощать к тех, кто пытался воспрепятствовать их появлению. Если вам близка гипотеза симуляции, то (учитывая объем идущей сейчас работы, направленной на подчинение или ограничение еще не существующего искусственного интеллекта), вероятно, вам стоит думать, что этот мир – как раз такая симуляция, созданная, чтобы истязать тех, кто не помог сверхразуму появиться раньше. А если вы работаете над созданием искусственного интеллекта, то, возможно, наши сверхразумные механические повелители будут благосклонны к вам.

Наши господа, рабы или партнеры?

Джон Маркофф

Ведущий автор отдела науки в New York Times; автор книги «Homo Roboticus? Люди и машины в поисках взаимопонимания» (Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots)[5]

Гегель сказал, что в отношениях между господином и рабом человеческий облик теряют оба. Эта идея взволновала широкий круг мыслителей, от Маркса до Бубера, не стоит забывать о ней и сегодня.

Хотя нет никаких доказательств, что вот-вот появятся машины, способные мыслить подобно человеку, неоспорим тот факт, что в мире, где все связано через интернет, скоро искусственный интеллект научится имитировать большую часть человеческой деятельности – и физической, и интеллектуальной. Так каковы же будут отношения между нами и нашими все более и более одаренным копиями?

Мы уже проводим немало времени, либо общаясь с другими людьми посредством компьютеров и сетей, либо напрямую взаимодействуя с «умными» машинами – в видеоиграх, в различных консалтинговых системах от так называемых FAQ-ботов, которые выдают текстовые ответы на набранные с клавиатуры вопросы, до человекоподобных программных аватаров. Станут ли эти аватары нашими рабами, помощниками, коллегами или в них будет понемногу от всех трех ипостасей? Или же, что выглядит куда более зловеще, они станут нашими господами?

Сама идея о том, чтобы размышлять о роботах и искусственном интеллекте с точки зрения общественных отношений может поначалу показаться неправильной. Однако мы склонны наделять свои машины человеческими качествами, даже когда их способности минимальны, и уж тем более будем делать это, если они станут независимыми. Компьютеры с функциями речевого взаимодействия уже сейчас слишком похожи на людей. Таким образом, разработчикам роботов будущего следует создавать товарищей, а не слуг, написать программу, действующую как музыкальный аккомпаниатор, а не как раб.

Если нам это не удастся, то может повториться исторический прецедент. Создав в умных электронных помощников, мы, вполне вероятно, столкнемся с той же проблемой, что и древние римляне, которые переложили собственную административную работу на рабов-греков и вскоре уже были не в состоянии думать самостоятельно.

Возможно, мы уже катимся по той же скользкой дорожке. Например, появляется все больше подтверждений, что, полагаясь на GPS в выборе маршрута и исправлении ошибок навигации, мы теряем способность к запоминанию направлений движения и пространственному мышлению – полезные для выживания навыки.

Этот пример указывает и на еще одну большую проблему: опасность передать контроль над повседневным решениями неким сложным алгоритмам.

Для нынешнего юного поколения мир перевернут вверх дном. Вместо того чтобы задействовать автоматы для рутинной работы и освободить время для серьезных мыслей, близких отношений и проявления собственной индивидуальности, творчества и свободы, молодежь обращается к смартфонам за советами. То, что началось как интернет-технологии для обмена тем, что нравится, по сути, стало разрастающейся массой алгоритмов, жаждущих новых данных и принимающих решения за нас.

Теперь уже интернет незаметно задает нам жизненные направления. Это могут быть мелочи вроде выбора лучшего в окрестностях корейского ресторана, причем выбора, осно-

[5] Маркофф Дж. Homo Roboticus? Люди и машины в поисках взаимопонимания. – М.: Альпина нон-фикшн, 2017.

ванного на все более качественном понимании Сетью ваших личных желаний и потребностей, или вещи более значимые, такие как сервисы для подбора супругов. Компьютер подберет вам не только ресторан, подарки или цветы, но и человека, с которым вы свяжете жизнь.

Основная мысль состоит в том, что разработчики программного обеспечения, специалисты по искусственному интеллекту, робототехники и хакеры, которые проектируют системы будущего, обладают властью менять общество.

Почти век назад Торстейн Веблен написал оказавшую серьезное влияние статью с критическим анализом промышленного мира начала XX века – она называлась «Инженеры и система ценообразования». Поскольку производственные технологии обрели силу и общественное значение, он полагал, что политическая власть перейдет к инженерам, чьи глубокие технологические знания станут основой для системы управления нарождающейся промышленной экономики. Все, конечно, вышло совсем не так. Веблен искал компромисс между марксизмом и капитализмом. Возможно, он опередил свое время, но его основная идея около тридцати лет спустя, на заре компьютерной эры, эхом отозвалась в работах Норберта Винера и, по-видимому, оказалась верной.

Возможно, Веблен не ошибся, а просто немного поторопился. В наше время инженеры, которые создают программы и роботов с элементами искусственного интеллекта, имеют огромное влияние на то, как именно мы их используем. По мере того как компьютерные системы все глубже проникают в повседневную жизнь, противоречия между дополненным интеллектом и искусственным интеллектом становятся все более заметными.

У Норберта Винера было ясное понимание того, какое значение имеют взаимоотношения между людьми и мыслящими машинами. Он видел преимущество автоматизации в устранении человека от рутинной работы, но также ясно сознавал и возможность порабощения человечества. Прошедшие с тех пор несколько десятилетий только обострили дихотомию, которую он обнаружил первым.

Дело в нас, в людях, и в том, какой мир мы создадим. Дело не в машинах, и неважно, насколько гениальными они станут.

Я, со своей стороны, не стану приветствовать ни роботов-повелителей, ни роботов-рабов.

Спроектированный интеллект

Пол Дэвис

Физик-теоретик, космолог, астробиолог; один из руководителей исследовательского центра BEYOND, Университет штата Аризона; ведущий исследователь Центра по взаимодействию физических наук и биологии раковых заболеваний; автор книги «Зловещая тишина: Одиноки ли мы во Вселенной» (The Eerie Silence: Renewing Our Search for Alien Intelligence)

Обсуждение искусственного интеллекта имеет отчетливый привкус 1950-х, и самое время прекратить уже использовать термин «искусственный». В действительности мы имеем в виду спроектированный интеллект (СИ). В просторечии такие слова, как «искусственный» и «машина», используются как антонимы слова «естественный» и содержат намеки на металлических роботов, электронные схемы и компьютеры, которым противопоставляются живые, пульсирующие, мыслящие биологические организмы. Сама идея о том, что у металлической штуковины, начиненной проводами, могут быть права или что она вдруг не подчинится людским законам, не просто жутка – она абсурдна. Но это совершенно не тот путь, которым идет СИ.

Очень скоро граница между искусственным и естественным исчезнет. Спроектированный разум будет все больше полагаться на синтетическую биологию и органические материалы, в которых выращенные из генетически модифицированных клеток нейронные сети станут произвольно самоорганизовываться в функциональные модули. На первых порах проектировщиками будут люди, но их очень скоро заменят более умные системы СИ, и начнется процесс безудержного усложнения. В отличие от ситуации с человеческими мозгами, которые довольно посредственно коммуницируют друг с другом, системы СИ будут связаны непосредственно и всесторонне, так, что понятие личного «я» перестанет существовать, а уровень познавательной деятельности поднимется до беспрецедентных высот. Возможно, что часть такой спроектированной нейронной сети будет включать в себя квантовые эффекты, приближаясь к идее Фрэнка Вильчека о «квантовом сознании». Такие сущности окажутся настолько далеки от сферы индивидуального человеческого мышления и сопутствующих ему квалиа[6], что почти все привычные вопросы, касающиеся возможностей и опасностей искусственного интеллекта, станут неактуальны.

Что же тогда будет с людьми? На пути дополненного человеческого разума, использующего схожую технологию, стоят только этические барьеры, и об этом уже давно размышляют философы-трансгуманисты. Генетически модифицированные люди с дополненными мозгами могли бы существенно развить и улучшить человеческие знания.

В таком случае у нас есть три варианта будущего, со своими этическими проблемами в каждом. В первом сценарии люди воздерживаются от усовершенствования самих себя и соглашаются отказаться от гегемонии в пользу СИ. Во втором сценарии люди, вместо того чтобы ограничивать себя, модифицируют собственные мозги (и тела), используя ту же технологию, что и при создании СИ, а впоследствии передают ему управление подобными усовершенствованиями – и тогда они становятся этакими сверхлюдьми, способными сосуществовать с СИ, но все же остающимися слабее него. Наконец, возможен и такой сценарий, при котором в определенный момент произойдет слияние СИ и дополненного человеческого интеллекта.

Если мы не одиноки во Вселенной, то нам следует ожидать контакта не с разумными существами плоти и крови, привычными нам по научно-фантастическим произведениям, а

[6] В философии этим термином обозначают любые ощутимые человеком явления, например цвет или вкус. – *Прим. ред.*

с возникшим много миллионов лет назад СИ, обладающим невообразимой мощью разума и решающим непостижимые для нас задачи.

Сверхразумный отшельник

Кевин Хэнд

Заместитель начальника отдела исследований Солнечной системы, Лаборатория реактивного движения НАСА, Калифорнийский технологический институт

Неизбежность появления мыслящих машин давно волнует тех из нас, кто смотрит в ночное небо и задается вопросом: живем ли мы во Вселенной, где кишит жизнь, или же она встречается чрезвычайно редко?

Проблема отлично сформулирована Энрико Ферми в виде вопроса: «Где они?» Если считать ориентиром нашу цивилизацию, разумные машины должны появиться в относительно краткосрочной перспективе (меньше нескольких тысяч лет после того, как были созданы компьютеры), а затем непосредственной задачей этих машин – так называемых зондов фон Неймана – станет распространение по другим звездным системам и быстрое воспроизводство. Таким образом, Галактика будет колонизирована в течение нескольких сотен миллионов лет, а это довольно быстро в сравнении с возрастом Вселенной (13,8 миллиарда лет) и даже Солнечной системы (4,6 миллиарда лет). Согласно парадоксу, сформулированному Ферми, если сверхразумные машины возникли где-нибудь в пределах Галактики, то они должны уже быть здесь. Так как мы их не наблюдаем, утверждают некоторые, технологически развитой жизни в Галактике еще нет.

Но совсем не обязательно сверхразумное существо должно испытывать то же эволюционное давление, которое заставляет нас исследовать окружающий мир (говоря «нас», я имею в виду непрочные водянистые кульки, именуемые людьми). Действительно ли исследование – это и биологический, и технологический императив? Будет ли у разумных машин мотивация для экспансии?

Основные причины, побуждающие нас заниматься исследованиями, следующие: свобода, любопытство и доступ к ресурсам. Из этих трех мотивов только последний может оказаться императивом для сверхразумного существа, поскольку первые два должны быть в значительной степени разрешены в процессе его становления. Доступ к ресурсам, конечно, сильный стимул, но совсем не обязательно для сверхразума «больше» всегда значит «лучше». В определенный момент материальных и энергетических ресурсов в пределах одной звездной системы должно быть достаточно на любые вычисления или симуляции. Воспроизводство, которое является разновидностью ресурсных потребностей, теряет актуальность для бессмертной машины, способной к самовосстановлению. Исследование ради стабильности в долгосрочной перспективе, конечно, придется принять во внимание: такие звезды, как наше Солнце, через несколько миллиардов лет могут прислать космическое уведомление о выселении. Правда, на то, чтобы найти недвижимость рядом с хорошим стабильным карликом класса М, не должно понадобиться слишком много времени, и нам остается только гадать, двинется ли сверхразум дальше после первого переселения. Будут ли у него какие-то причины, чтобы продолжать странствия?

Желание проверить какие-либо собственные компьютерные модели или теории о строении Вселенной могли бы побудить сверхразумное существо к исследованиям. Но такие эксперименты не требуют колонизации. Например, научные программы, проводимые НАСА, не очень сильно мотивированы необходимостью колонизации. Нет нужды рисковать жизнями людей, для того чтобы исследовать океан Европы, спутника Юпитера (хотя там, наверное, есть на что посмотреть). Мне кажется, что сверхразум для исследования черной дыры и проверки своих расчетов просто пошлет флот из роботов навстречу ненапрасной, хоть и

неминуемой гибели. Любопытство[7] в случае со сверхразумным существом запросто может принять форму робота под управлением робота.

Отмечу, что интеллект и исследование физического мира не всегда были тесно связаны и в нашей цивилизации. Возможно, из-за с стремления к самосохранению или просто из-за желания сосредоточить мысль на чем-то конкретном один и тот же индивид редко пытался раздвинуть и границы разума, и границы физики (Дарвин, вероятно, стал одним из редких исключений). Почему то же самое не может быть справедливо для разумных машин?

Не исключено, что их удел состоит в том, чтобы вращаться вокруг нежаркой, стабильной звезды класса М, год за годом прогоняя через себя симуляции окружающего мира из чистого любопытства и удовольствия от получения верных результатов. Эти сверхразумные существа могли бы быть космической версией мудреца-отшельника, живущего в лесной хижине, которому вполне хватает для счастья собственных размышлений и самопознания.

[7] Игра слов: Curiosity (англ. «любопытство») – название марсохода. – *Прим. пер.*

Это будет дикая гонка

Джон Мазер

Главный астрофизик Лаборатории наблюдательной космологии, Центр космических полетов им. Годдарда, НАСА; автор книги «Самый первый свет: Подлинная история научного путешествия к началу Вселенной» (The Very First Light: The True Inside Story of the Scientific Journey Back to the Dawn of the Universe)

Мыслящие машины эволюционируют точно так же, как описанные Дарвином живые (и мыслящие) биологические виды, – посредством конкуренции, сотрудничества, борьбы за выживание и воспроизводства. Машины становятся интереснее по мере того, как они учатся понимать физические объекты и управлять ими либо непосредственно, либо при участии людей.

Мы пока не обнаружили такого закона природы, который препятствовал бы появлению настоящего универсального искусственного интеллекта, так что я думаю, что это произойдет, и довольно скоро, учитывая триллионы долларов, что люди инвестируют в электронные аппаратные средства, а также те триллионы, которые заработают потенциальные победители. Эксперты говорят, что мы недостаточно хорошо понимаем, что такое интеллект, чтобы его построить, и тут согласен, но набор из сорока шести хромосом этого тоже не понимает и тем не менее управляет формированием известного нам самопрограммируемого биокомпьютера. Другие эксперты говорят, что закон Мура скоро перестанет работать, и мы не сможем позволить себе совершенствовать аппаратные средства в прежнем темпе; даже если это и так, «скоро» – понятие растяжимое.

Я пришел к выводу, что мы уже поддерживаем эволюцию мощного искусственного интеллекта, а он, в свою очередь, повлияет на развитие привычных нам могущественных сил: бизнеса, индустрии развлечений, медицины, государственной безопасности, производства оружия, власти на всех уровнях, преступности, транспорта, горнодобывающей промышленности, производства, торговли, секса – да чего угодно!

Я думаю, что результаты нам не понравятся. А они могут появиться очень быстро, настолько быстро, что великие империи падут и на их месте вырастут новые, а у людей будет самый минимум времени, чтобы приспособиться к жизни в новой действительности. Я не знаю, окажется ли кто-нибудь достаточно умным и одаренным для того, чтобы сохранить власть над этим джинном, потому что контролировать, возможно, придется не только машины, но и людей, дорвавшихся до новых технологий и имеющих злые намерения.

Что случится, когда умные роботы станут справляться с большинством наших рутинных дел? Кто будет их строить, кто будет ими владеть и кто потеряет работу? Будут ли роботы распространены только в странах с развитой экономикой или начнется коммерческая экспансия высоких технологий в остальные части мира? Будут ли они достаточно дешевыми, чтобы вытеснить всех фермеров с полей? Будет ли каждая машина обладать собственной личностью и не придется ли нам думать, в какой детский сад ее отправить, в какую школу или колледж? Будут ли роботы конкурировать друг с другом за рабочие места? Станут ли они величайшими гиперсоциальными хищниками, вытеснив людей и сделав их гражданами второго сорта или даже чем-то меньшим? Будут ли они заботиться об окружающей среде? Будут ли обладать чувством ответственности, и если да, то будут ли наделены им изначально или же разовьют его самостоятельно? Нет никакой гарантии, что они станут следовать трем законам робототехники Айзека Азимова.

С другой стороны, я ученый, и мне интересно, как можно применить искусственный интеллект к исследованию новых наук и технологий. Преимущества для космических исследований очевидны: машинам не нужно дышать, они могут выдерживать экстремальные тем-

пературные и радиационные условия. Так что им будет куда легче, чем нам, колонизировать Марс. Они смогут отправиться во внешнюю область Солнечной системы, имея больше шансов выйти оттуда на связь, чем нынешние автоматические космические аппараты, а когда-нибудь смогут и полететь к звездам – если захотят.

Так же дело обстоит и с морскими глубинами. У нас уже есть тяжелая техника, работающая на донных буровых установках, но про морское дно мы все еще почти ничего не знаем, а ценность находящихся под водой минеральных и энергетических ресурсов неисчислима. Не исключено, что когда-нибудь случится война роботов в глубинах океана.

Мыслящие машины могут оказаться похожими на нас, с таким же желанием все исследовать, но могут быть и совсем другими. Зачем мне или роботу отправляться в путешествие длиною в тысячу лет к другой звезде через темноту космоса, не имея ни связи с товарищами, ни надежд на спасение, если что-то пойдет не так? Кто-то на такое согласится, кто-то нет. Возможно, мыслящие машины окажутся во многом похожими на свои биологические прототипы.

Это будет дикая гонка далеко за пределами наших самых смелых фантазий или кошмаров. Нет другого способа построить галактическую цивилизацию, за исключением варп-двигателя[8], и, возможно, кроме нас, никто во всем Млечном Пути не способен на это. Но не факт, что мы переживем столкновение с иным разумом, который сами же и создали.

[8] Несуществующая, но теоретически возможная технология, позволяющая перемещаться быстрее скорости света. – *Прим. ред.*

Этой штукой кто-нибудь управляет?

Дэвид Кристиан

Профессор истории, Университет Маккуори (Сидней); автор, совместно с Синтией Стокс-Браун и Крэйгом Бенджамином, книги «Карты времени: Между ничем и всем» (Maps of Time: Between Nothing and Everything)

Вселенная существует 13,8 миллиарда лет, люди – 200 000 лет, этот временной отрезок – всего 1/69 000 возраста Вселенной. Менее 100 лет назад люди создали машины, способные самостоятельно выполнять затейливые вычисления. Чтобы увидеть мыслящие машины в правильном свете, нам надо подумать об истории мышления.

Мышление как таковое и мышление все более и более сложное – это явления, принадлежащие большой истории, истории того, как наша Вселенная создавала все более и более сложные сети вещей, скрепленных между собой энергией, причем каждая такая система обладала новыми эмерджентными свойствами[9]. Звезды – структурированные облака протонов, скрепленные энергией термоядерного синтеза. Когда разрушались большие звезды и образовывались сверхновые, возникали не существовавшие прежде типы атомов; электромагнетизм стягивал атомы в скопления льда и минеральной пыли, а сила тяжести спрессовывала молекулы в большие системы, которые мы называем планетами. Мышление возникло в рамках еще более сложных систем, сформированных живыми организмами. В отличие от объектов, которые существуют в состоянии равновесия, таких как звезды или кристаллы, биологическим существам приходится выживать в нестабильной окружающей среде. Они плывут через постоянно меняющуюся кислотность, температуру, давление и т. п. Таким образом, они должны все время приспосабливаться. Мы называем это гомеостазом – именно он создает ощущение, что у живых организмов есть цель и способность делать выбор. Короче говоря, они как будто думают. Они могут выбирать варианты действий, оставляющие достаточное количество энергии, чтобы организм продолжал функционировать. Так что их выбор отнюдь не случаен. Наоборот, естественный отбор – это процесс, в котором большую часть времени большинство организмов будут совершать действия, увеличивающие их шансы контролировать энергию и ресурсы, необходимые им для выживания и воспроизводства.

Нейроны – это такие занятные клетки, которые хорошо умеют принимать решения. Когда они образовывают сети, получается мозг. Система из нескольких нейронов уже способна принимать разные решения, и число таковых увеличивается экспоненциально по мере расширения сети нейронов. Решения, которые принимает мозг, чтобы реагировать на вызовы окружающей среды, бывают весьма изощренными. По мере того как организмы становились более сложными, клетки объединялись в сети и создавали величественные органические конструкции – биологические эквиваленты Эмпайр-стейт-билдинг или Бурдж-Халифа. Нейроны в мозгах этих организмов создавали еще более замысловатые сети, чтобы получить возможность управлять неповоротливыми телами с исключительной точностью и изобретательностью и обеспечить им выживание и воспроизводство. Прежде всего мозгам надо было обеспечить своим организмам возможность подключиться к потокам энергии, производимой Солнцем и пригодной для потребления посредством фотосинтеза.

Люди добавили еще один уровень организации сети, когда язык связал между собой мозги из разных регионов и поколений, создав обширные региональные интеллектуальные сети. Это называется коллективным обучением. Его мощь увеличивалась по мере того,

[9] Эмерджентными называются свойства, присущие системе в целом, но не присущие ни одному из ее элементов по отдельности. – *Прим. ред.*

как росла эффективность сотрудничества людей в рамках все более и более многочисленных сообществ, которые учились использовать все более значительные потоки энергии. За последние два века эти сети стали глобальными, а мы научились использовать огромные запасы окаменевшего солнечного света, хранящиеся под землей миллионы лет. Вот почему наше воздействие на биосферу в ходе антропоцена оказалось настолько значительным.

Коллективное обучение также породило средства, расширяющие мышление, – от устных преданий до письменности, книгопечатания и науки. Каждое из них усиливало мощь поразительной мыслящей машины, составленной из сети человеческих мозгов. Однако за последний век в результате сочетания ископаемого топлива и вычислительных машин она развилась гораздо больше, чем за всю предыдущую историю. Поскольку компьютеры за последние 30 лет выстроили собственные сети, они многократно увеличили коллективную мощь человечества.

Самая могущественная из известных нам на сегодняшний день разумных машин собрана из миллиардов человеческих мозгов, каждый из которых представляет собой огромную сеть нейронов. Все эти мыслящие единицы сперва были объединены в пространстве и времени, а теперь еще и усилены миллионами объединенных в сети компьютеров.

Этой штукой кто-нибудь управляет? Что-то удерживает вместе ее элементы? Если да, то кому она служит и чего хочет? А если ею никто не управляет, то неужели колосс современного общества вообще неуправляем? Жутковато, не правда ли? Меня больше всего волнует даже не то, о чем именно думает эта огромная машина, а то, является ли ее мышление хоть сколько-нибудь последовательным. Не будут ли отдельные части тянуть ее в разных направлениях, не сломают ли они ее в конце концов и не обернется ли это катастрофой для детей наших детей?

Свидетель Вселенной

Тимо Ханней

Исполнительный директор в Digital Science, Macmillan Science and Education; один из основателей Sci Foo

Если отталкиваться от одного из определений слова «мыслить» – «собирать, обрабатывать информацию и действовать на ее основе», можно сделать вывод, что Землю наводнили основанные на кремнии мыслящие машины. Все – от термостатов до телефонов – устройства, которые привносят в нашу повседневную жизнь удобство и радость, теперь наделены настолько впечатляющим интеллектом, что мы уже привычно говорим о них, без тени иронии, что они умные. Наши самолеты, поезда, а теперь еще и автомобили становятся практически полностью самостоятельными, и, наверное, недалек тот день, когда они смогут избавиться от одной из главных причин поломок, задержек и аварий – людей-операторов.

Мало того, навыки этих машин быстро развиваются, что обусловлено доступом к большим, чем когда-либо, объемам данных и вычислительным мощностям вместе со стремительным (хоть и не всегда понятным) улучшением алгоритмов. После десятилетий несбыточных обещаний специалисты в технических науках вдруг обнаружили, что их творения способны на сверхчеловеческий уровень производительности в таких ранее сложных для них областях, как распознавание голоса, рукописного текста и изображений, не говоря уже о тестах на общую эрудицию. Модель развития по схеме «стой – иди» была настолько необыкновенной, что если бы кого-нибудь переместили во времени из 2010-го в 2015-й, то он был бы поражен нынешним состоянием техники куда больше, чем путешественник во времени, прибывший из середины XX века или более ранней эпохи.

Но если индустрия искусственного интеллекта уже не шутка, то не превратилась ли она во кое-что похуже – в плохой фильм ужасов?

Машины теперь знают намного больше, чем любой из нас, лучше справляются со многими задачами, нисколько не уставая, – так чего им стоит перехватить инициативу и стать нашими хозяевами? И не наступит ли время, когда самые продвинутые из наших творений породят еще более умные машины, которые мы вообще не сможем ни понимать, ни контролировать?

Возможно, такие риски и стоит учитывать. Но если нас и ждут эти проблемы, то лишь в отдаленном будущем. Искусственный интеллект хоть и выглядит впечатляюще в определенных областях, все еще остается достаточно ограниченным и негибким. Самая замечательная особенность биологического разума состоит не в его мощности, а в удивительной универсальности – от отвлеченного полета фантазии до высочайшего совершенства в физическом развитии, иначе говоря, от Дворжака до Джоковича[10].

По этой причине люди и машины скорее продолжат дополнять друг друга, чем станут конкурировать, а самые сложные задачи – навигация в физическом мире, лечение болезней, сражение с врагом на поле битвы – будут лучше всего выполняться в ходе совместной работы углерода и кремния. Пока что наибольшую опасность для человечества представляют сами люди. Чтобы оказаться реальной угрозой, машинам сначала надо стать более похожими на нас, а сейчас почти никто не пытается создать ничего подобного – гораздо проще и интереснее делать людей.

И все же если мы действительно рассматриваем долгосрочные перспективы, то существует серьезная причина сделать машины более похожими на нас по одному важнейшему – и пока что совершенно не учитываемому – критерию. Если отталкиваться от другого опре-

[10] Имеются в виду композитор Антонин Дворжак (1841–1904) и теннисист Новак Джокович (род. 1987). – *Прим. ред.*

деления слова «мыслить», то машинам это вообще не свойственно, потому что ни одна из них не обладает самосознанием. Точнее говоря, мы не можем знать или даже с достоверностью предполагать, есть ли вероятность появления сознания у основанного на кремнии интеллекта, хотя большинство из нас считает, что нет. Есть три объяснения, почему создание действительно обладающего сознанием искусственного интеллекта – это хорошо. Во-первых, так мы получим свидетельство, что у нас есть в целом приемлемая теория насчет того, что необходимо для получения субъективного опыта. Во-вторых, сам факт, что одно мыслящее существо намеренно (позволено ли мне будет сказать «сознательно»?) конструирует другую форму разума, займет достойное место в ряду важнейших событий истории.

В-третьих, Вселенная без интеллекта, обладающего самосознанием и способного ее наблюдать, в конечном счете бессмысленна. Мы не знаем, есть ли где-то там другие существа, но можем не сомневаться, что рано или поздно исчезнем. Сознательный искусственный интеллект мог бы пережить наш неизбежный конец и даже грядущую гибель всего живого на Земле, когда Солнце вздуется и превратится в красного гиганта. Задача такой машины будет состоять не только в том, чтобы мыслить, но также и в том – и это куда важнее, – чтобы не дать угаснуть мерцающему пламени сознания, стать свидетелем Вселенной и чувствовать ее чудо.

Давайте подготовимся!

Макс Тегмарк

Физик, космолог, Массачусетский технологический институт; директор по научным исследованиям в Институте основополагающих вопросов; один из основателей Института основополагающих вопросов; автор книги «Наша математическая вселенная» (Our Mathematical Universe)

Самый интересный для меня вопрос об искусственном интеллекте состоит не в том, что мы думаем о нем, а в том, что мы с ним делаем. В связи с этим в нашем недавно созданном Институте основополагающих вопросов мы привлекаем к обсуждению будущего отрасли многих выдающихся мировых специалистов по искусственному интеллекту. Вместе с ведущими экономистами, правоведами и другими экспертами мы изучаем все классические вопросы, такие как:

• Что произойдет с людьми, если машины постепенно заменят их на рынке труда?

• Когда, если это вообще произойдет, машины превзойдут людей во всех интеллектуальных задачах?

• Что будет потом? Произойдет ли взрывное развитие машинного разума, которое оставит нас далеко позади, и если да, то какая роль останется людям после этого (и останется ли вообще хоть какая-нибудь)?

Необходимо провести огромное количество исследований прямо сейчас, чтобы не только обеспечить функциональность систем искусственного интеллекта, но также гарантировать их надежность, полезность и управляемость.

Как и в случае с любой новой технологией, здесь вполне естественно сначала будет сосредоточить внимание на том, чтобы она заработала. Но как только забрезжит успех, придется задуматься о воздействии технологии на общество и выяснить, как нам беспроблемно пользоваться преимуществами искусственного интеллекта и избегать опасностей. Вот почему, научившись добывать огонь, мы разработали огнетушители и нормы пожарной безопасности. Для более современных технологий, таких как ядерная энергия, синтетическая биология и искусственный интеллект, оптимизация социального воздействия становится еще более важной задачей. Короче говоря, наша технологическая мощь должна соответствовать нашей мудрости в плане ее применения.

К сожалению, необходимый нам здравый подход к выбору задач для научных исследований, практически полностью глушит какофония невежественных суждений, охвативших блогосферу. Позвольте вкратце перечислить самые громкие из них.

1. Паникерство: страх увеличивает доходы от рекламы и рейтинг Нильсена[11], а многие журналисты, кажется, не способны написать статью про искусственный интеллект без картинки с роботом, несущим оружие.

2. «Это невозможно»: будучи физиком, я знаю, что мой мозг состоит из кварков и электронов, организованных и действующих подобно мощному компьютеру, и что нет такого закона физики, который препятствовал бы тому, чтобы мы построили еще более разумные сгустки кварков.

3. «Это произойдет не при нашей жизни»: мы не знаем, каковы шансы, что машины достигнут человеческого уровня во всех когнитивных задачах при нашей жизни, но большинство исследователей, работающих в области искусственного интеллекта, на прошедшей недавно конференции высказывались в пользу того, что вероятность этого выше 50 процен-

[11] Рейтинг популярности телевизионной программы. – *Прим. ред.*

тов, таким образом, было бы глупо с нашей стороны отбрасывать такую возможность, считая ее просто научной фантастикой.

4. «Машины не могут контролировать людей»: мы способны управлять тиграми не потому, что мы сильнее, а потому, что умнее, так что если мы сдадим позиции самых умных на планете, то рискуем утратить свободу.

5. «У машин нет целей»: многие системы искусственного интеллекта запрограммированы таким образом, чтобы иметь цели и максимально эффективно их достигать.

6. «Искусственный интеллект сам по себе не злонамерен»: верно, но его цели могут однажды столкнуться с вашими. Люди обычно не питают ненависти к муравьям, но если б мы захотели построить плотину ГЭС, а на ее месте оказался бы муравейник, то муравьи столкнулись бы с проблемами.

7. «Люди достойны того, чтобы их заменили машины»: спросите любого родителя, как бы он отнесся к идее заменить своего ребенка машиной и хотел бы он иметь право голоса в этом вопросе.

8. «Те, кто беспокоится по поводу искусственного интеллекта, не понимают, как работают компьютеры»: это утверждение прозвучало на вышеупомянутой конференции, и собравшиеся там специалисты по ИИ сильно смеялись.

Давайте не будем позволять шумихе вокруг этих «красных селедок»[12] отвлекать нас от главной проблемы. Воздействие искусственного интеллекта на человечество неуклонно растет, и, чтобы гарантировать, что это воздействие останется положительным, необходимы общие усилия по исследованию ряда сложнейших проблем – междисциплинарных, касающихся и общества, и искусственного интеллекта, а потому требующих сотрудничества исследователей во многих областях. Начинать работу над этими проблемами нужно уже сейчас.

Сперва мы, люди, придумали, как воспроизвести некоторые природные процессы с помощью машин и получили рукотворный ветер, молнии и лошадиные силы. Постепенно мы поняли, что наши организмы также являются машинами, а после открытия нервных клеток граница между телом и разумом начала размываться. Тогда мы стали строить машины, способные превзойти нас не только в силе, но и в интеллекте. Так не станем ли мы анахронизмом, выяснив, кто мы на самом деле?

Появление по-настоящему мыслящих машин будет самым важным событием в истории человечества. Станет ли это лучшим или худшим из всего случившегося с человечеством, зависит от того, как мы подготовимся, и сейчас самое время начинать. Не нужно быть сверхразумным искусственным интеллектом, чтобы понять: двигаться навстречу величайшему событию в истории человечества и не готовиться к этому – просто глупо.

[12] Английская идиома, обозначающая намеренную дезинформацию с целью отвлечь чье-либо внимание. Копченую («красную») сельдь, обладающую сильным запахом, традиционно использовали для натаскивания охотничьих собак. – *Прим. ред.*

Вопросы для расширенного теста Тьюринга

Томазо Поджо

Профессор отделения когнитивистики и наук о мозге, директор Центра изучения мозга, разума и машин, Массачусетский технологический институт

В последние несколько месяцев мы наблюдали охватывающее все более и более широкие слои общества обсуждение вопросов, касающихся опасности искусственного интеллекта, и в частности сильного искусственного интеллекта (СИИ). Некоторые – и в их числе физик Стивен Хокинг – назвали ИИ главной угрозой существованию человечества, а такие фильмы, как «Она» (Спайк Джонз, 2013) и «Превосходство» (Уолли Пфистер, 2014), подкрепили общие настроения. И даже разумные замечания экспертов в данной области – среди которых были Род Брукс и Орен Этциони – не то чтобы смогли направить обсуждение в позитивное русло.

Я утверждаю, что изучать то, как мы думаем о мыслящих машинах и их создании, полезно для общества. Я призываю к исследованиям, которые объединят нейронауки, когнитивистику, информатику и собственно разработку искусственного интеллекта. Понимание природы мышления и репликация его в машинах идут рука об руку с осознанием того, как выполняет свою работу мозг.

Сближение математики, нейробиологии и технологических разработок создает новые возможности междисциплинарного взаимодействия. Мечта о том, чтобы понять, как работает мозг, не нова, и все же – как видно из дебатов вокруг ИИ – сейчас подходящее время для того, чтобы сделать шаг к этой мечте. Мы видим зарождение новой области – науки и техники интеллекта, в которой нам предстоит достичь фундаментальных, очень важных для человечества результатов. Мы должны продолжать исследования, а не сворачивать их.

Проблема интеллекта – что это такое, как он порождается человеческим мозгом и как воспроизвести его в машинах – одна из наиболее важных проблем науки и техники, наряду с проблемой происхождения Вселенной и природы пространства и времени. Она может оказаться важнейшей, потому что обладает эффектом мультипликатора: практически любое достижение, способное сделать нас умнее непосредственно или с помощью мыслящих машин, позволит добиться прогресса в решении других важных проблем науки и техники.

Исследования в области интеллекта в конечном счете приведут к революционным изменениям в образовании и обучении. Системы, способные выявлять влияние культуры на мышление, позволили бы нам свести на нет социальные конфликты. Умные машины помогли бы ученым и инженерам в решении наиболее острых технических проблем. Мы могли бы лучше разобраться в психических заболеваниях, а значит, найти лучшие способы их лечения. Таким образом, исследования в области интеллекта должны помочь нам понять, как работает человеческий разум, научиться строить более разумные машины и улучшить механизмы выработки коллективных решений. Эти достижения будут иметь решающее значение для нашего процветания, образования, здоровья и общественной безопасности. Сейчас, я повторяю, самое время для того, чтобы значительно расширить исследования в области интеллекта, а не сворачивать их.

Нас часто вводят в заблуждение общие, расплывчатые по значению, но часто повторяемые слова. Никто пока не смог дать точное, верифицируемое определение, что такое интеллект и мышление в целом. Единственная известная мне практически применимая характеристика, хоть и довольно ограниченная, принадлежит Алану Тьюрингу. В своем тесте он дал рабочее определение особой формы мышления – человеческого мышления.

Давайте рассматривать наше мышление так, как его характеризует тест Тьюринга. Становится очевидно, что у человеческого мышления много аспектов. Рассмотрим, например, тест Тьюринга на визуальный интеллект, то есть вопросы об изображении, варьирующие от «Что изображено на картинке?» до «Кто изображен на картинке?» и «Что делает этот человек?» и даже «Что эта девочка думает об этом мальчике?» и т. д. Сейчас благодаря последним достижениям нейронаук нам уже известно, что для ответа на такие вопросы требуются различные компетенции и умения, часто не зависящие друг от друга, а также относящиеся к разным отделам мозга. Внешне похожие вопросы, касающиеся распознавания предметов и лиц («Что там?» в противоположность «Кто там?»), соотносятся с разными частями зрительной зоны коры головного мозга. Слово «интеллект» в этом контексте может быть истолковано так же неверно, как слово «жизнь»: в течение первой половины прошлого века популярные научные журналы обычно писали о проблеме жизни так, будто существует некий единый жизненный субстрат, который лишь надо открыть.

Сегодня говорить о проблеме жизни смешно: биология – это наука, работающая со множеством серьезных проблем, а не с какой-то единственной. Интеллект – одно слово, скрывающее множество проблем, решение которых достойно множества Нобелевских премий. Это связано с точкой зрения Марвина Минского на проблему мышления, заключенную в его афористичной формуле «Общество разума». Так же обстоит дело и с настоящим тестом Тьюринга: это широкий круг вопросов, призванных исследовать главные аспекты человеческого мышления. Вот почему мы с коллегами занимаемся разработкой теоретических основ для неисчерпывающего набора вопросов расширенного теста Тьюринга, чтобы оценить научные достижения в нашей области. Слово «вопросы» во множественном числе подчеркивает, что мы говорим о большом числе различных интеллектуальных способностей, которые еще предстоит описать и, возможно, воспроизвести в искусственном интеллекте: базовое визуальное распознавание объектов, идентификация лиц, оценка эмоций, социальный интеллект, язык и многое другое. Расширенный тест Тьюринга делает особый упор на то, что количественная модель должна сопоставлять человеческое поведение и человеческую физиологию – разум и мозг. Таким образом, наши условия лежат далеко за пределами исходного теста Тьюринга; для их понимания и разработки соответствующих интеллектуальных технологий необходима целая научная область.

Следует ли нам бояться мыслящих машин?

Так как интеллект – это набор решений не зависящих друг от друга проблем, особых причин бояться внезапного появления сверхразумной машины нет, хотя всегда лучше перестраховаться. Конечно, любая технология из тех, что появляются сейчас или появятся через некоторое время, чтобы решать те или иные проблемы исследования интеллекта, вероятно, будет сильна сама по себе, а потому потенциально опасна в случае злоупотребления (впрочем, это касается большинства технологий).

Таким образом, здесь, как и в других областях науки, необходимы соответствующие меры безопасности и этические нормы. Кроме того, вероятно, понадобится постоянный мониторинг (возможно, силами независимой межгосударственной организации) супралинейного риска, создаваемого комбинациями постоянно появляющихся интеллектуальных технологий. Впрочем, подводя итог, я отмечу, что не только не боюсь разумных машин, но и считаю их зарождение и эволюцию одним из самых захватывающих, интересных и позитивных событий в истории человеческой мысли.

Эпохальное для человека событие

Памела Маккордак

Автор книги «Машины, которые думают» (Machines Who Think) и «Универсальная машина» (The Universal Machine); автор, совместно с Эдвардом Фейгенбаумом, книги «Пятое поколение: Искусственный интеллект и японский компьютерный вызов миру» (The Fifth Generation: Artificial Intelligence & Japan's Computer Challenge to the World)

Более пятидесяти лет я наблюдаю, насколько быстро меняется общественное мнение по поводу искусственного интеллекта. «Он невозможен, это нереализуемо». «Это важно / это не важно». «Это шутка». «Не может быть сильного искусственного интеллекта, только слабый». «Он уничтожит людей как вид». Такого рода крайности в последнее время уступили место признанию факта, что искусственный интеллект – это эпохальное научное, технологическое и социальное явление. Мы создали новый разум, который будет существовать наряду с нашим. Если мы станем обращаться с ним мудро, он принесет огромную пользу как всему человечеству, так и каждому из нас.

В качестве одного из вариантов будущего для искусственного интеллекта можно представить этакого Дживса, мудрого и терпеливого по сравнению с куда менее смышленым Берти Вустером, то есть с нами («Дживс, вы гений!» – «Спасибо, сэр. Мы стараемся»). Такой сценарий сосуществования возможен и даже желателен, помощь нам пригодится. Шахматы предлагают подобную модель. Гроссмейстеры Гарри Каспаров и Ханс Берлинер заявляли публично, что шахматные программы находят такие ходы, которых они сами не придумали бы, и учат игроков-людей новым трюкам. Deep Blue обыграл Каспарова, в то время как он считался одним из сильнейших шахматистов мира, при этом сам гроссмейстер и большинство экспертов полагают, что куда лучше в шахматы играют команды из людей и машин. Это ли не модель наших будущих взаимоотношений с искусственным интеллектом? Или, может, это лишь временный вариант, пока машины не овладеют и некоторыми сугубо человеческими навыками? Неизвестно. По скорости, широте и глубине мышления искусственный интеллект, вероятно, превзойдет человеческий – и уже много где превзошел.

Ни одна новая серьезная наука или технология не обходятся без недостатков, иногда даже опасных. Обнаружить их, оценить и принять необходимые меры – это задача громадных масштабов. Ею уже занимаются специалисты – философы, этики, правоведы и другие люди, обученные исследовать ценности за пределами простых интуитивных реакций, – в проекте под названием AI100, реализуемом на базе Стэнфордского университета. Никто не рассчитывает получить простые, окончательные ответы, понятно, что работа предстоит большая и продолжительная. Финансирование на ближайшие сто лет взял на себя один из ведущих специалистов в области искусственного интеллекта, Эрик Хорвиц, который вместе со своей женой Мэри и стал инициатором этого беспрецедентного исследования.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, купив полную легальную версию на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.

Комментарии

1.

Сет Ллойд. Программируя Вселенную. – М.: Альпина нон-фикшн, 2014).

2.

Steven Pinker, The Better Angels of Our Nature (New York: Viking Penguin, 2011). [Стивен Пинкер. Добрые ангелы человеческой природы]