

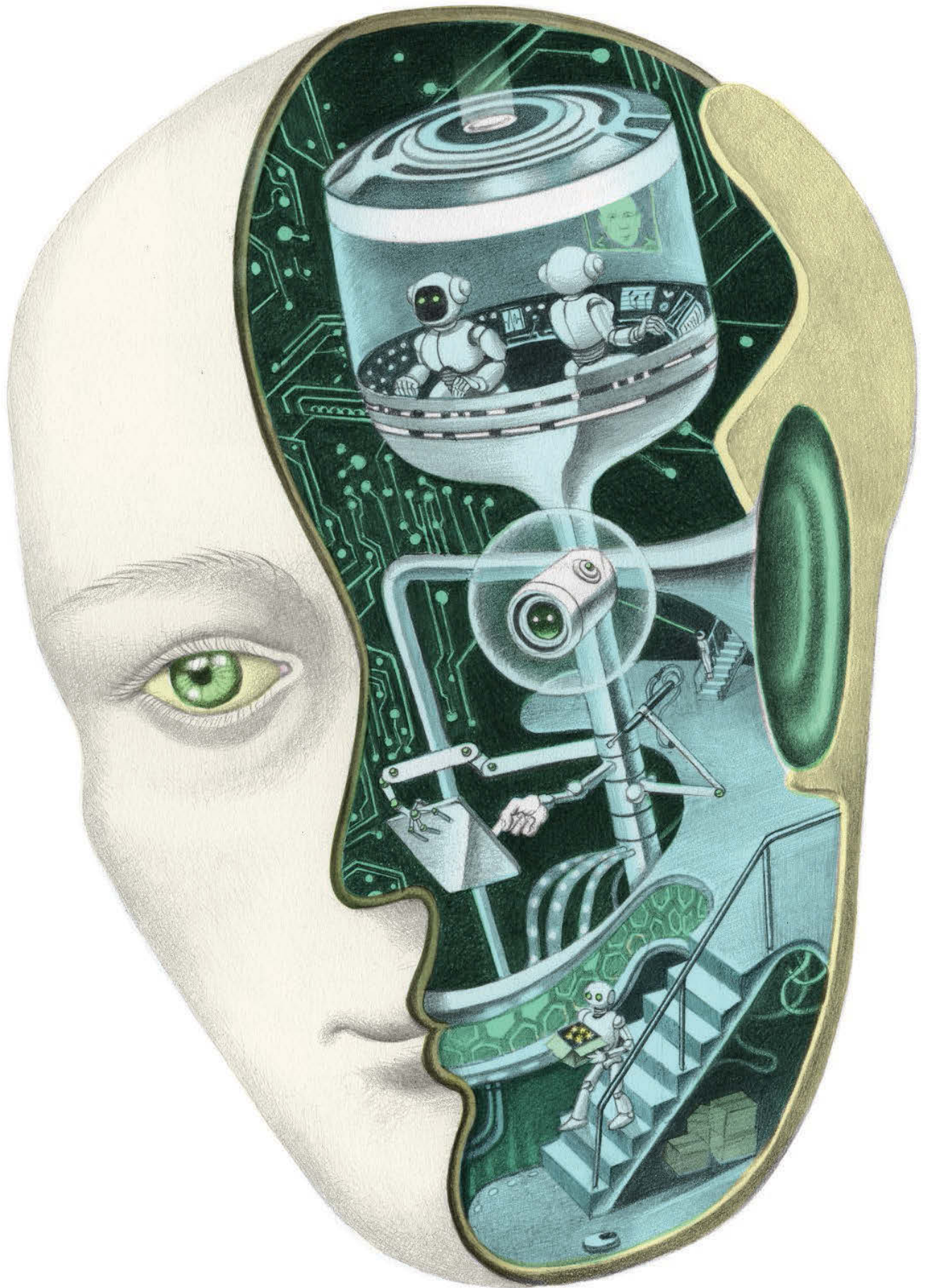
ЧАСТЬ III

Время
нас

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ — НАШ ПОМОЩНИК, А НЕ КОНТРОЛЕР

**НАШИ
ЦИФРОВЫЕ
ДВОЙНИКИ**

Педро Домингос



Ч

еловек — единственное живое существо на Земле, способное создавать машины. Но создавая их, он выходит в своих возможностях за рамки биологических ограничений. Благодаря инструментам и приспособлениям его руки могут делать гораздо больше, чем заложено в них природой. Сконструированные человеком автомобили мчат его со скоростью, недостижимой ранее, самолеты поднимают в воздух и переносят на огромные расстояния. Компьютеры «расширяют» наш головной мозг и увеличивают объем памяти, а смартфоны по-новому организуют повседневную жизнь. Разрабатываемые нами сегодня технологии способны к саморазвитию благодаря заложенной в них способности к использованию и переработке информации. Но не приведет ли это к тому, что умные машины попросту вытеснят нас? А может быть, они расширят наши возможности, доведя человеческие качества до совершенства?

Обучаемые машины появились в 1950-х гг. в результате пионерских работ Фрэнка Розенблатта (Frank Rosenblatt), который создал электронный нейрон, умеющий распознавать цифры, и Артура Сэмюэла (Arthur Samuel), снабдившего компьютер программой, следуя которой он играет с самим собой до тех пор, пока не победит некоего реального соперника. Но только за последние десять лет в этой области произошел настоящий прорыв: появились беспилотные автомобили, виртуальные помощники, которые понимают отдаваемые им приказы, и бесчисленное множество других приложений.

Каждый год разрабатываются сотни новых алгоритмов, задающих компьютеру те или иные последовательности действий. Однако принципиальная особенность обучаемых машин состоит в том, что вместо их детального программирования мы ставим перед ними глобальную цель, например: «Научись играть в шахматы». Как и человек, они начинают с приобретения «жизненного опыта». Обучающие алгоритмы распадаются на пять категорий, каждая из которых восходит к конкретной научной области. Один из примеров — обучение компьютера имитации естественного отбора через последовательное изменение алгоритмов.

ОСНОВНЫЕ ПОЛОЖЕНИЯ

- Создание искусственного интеллекта можно рассматривать как часть эволюции человека. На следующем ее этапе нам понадобится так называемый универсальный алгоритм. Он объединит в себе пять основных способов, какими нынешние обучаемые машины совершенствуют друг друга.
- Любая технология расширяет возможности человека. Машины не обладают свободой воли, цели и задачи задаем мы. Нас должно беспокоить не превосходство роботов над нами, а неправильное использование технологий.
- Наиболее вероятный сценарий развития ИИ на ближайшее будущее — совершенствование наших «цифровых двойников», виртуальных моделей, способных взаимодействовать с бесчисленным числом других. Все это должно помочь нам делать ежедневно множество осознанных выборов с невероятной скоростью.

ОБ АВТОРЕ

Педро Домингос (Pedro Domingos) — профессор компьютерных наук из Вашингтонского университета, автор книги «Верховный алгоритм» (*The Master Algorithm*, 2015). Член Ассоциации по продвижению искусственного интеллекта (AAAI). Живет близ Сиэтла.



В лаборатории креативных машин в Колумбийском университете пытаются научить примитивных роботов ползать или летать, для чего периодически комбинируют и видоизменяют 3D-печать для каждой следующей генерации. Начав со случайно собранных ботов, только-только научившихся перемещаться, конструкторы получали через тысячи генераций конструкции «робот-паук» и «робот-стрекоза».

Как известно, эволюция — медленный процесс. Углубленное обучение, наиболее популярная сегодня парадигма, инспирирует работу головного мозга. Мы начинаем с простейшей математической модели, описывающей работу нейрона, затем создаем сеть из тысяч миллионов элементов и обучаем ее постепенному упрочнению связей между нейронами, которые активизируются одновременно, как только получают сигнал. Такие нейронные сети способны распознавать лица, понимать человеческую речь и переводить тексты с одного языка на другой с беспрецедентной точностью. Машинное обучение начинает проникать в область психологии. Подобно нам, людям, аналоговые алгоритмы способны решать новые задачи, отыскивая сходные ситуации в памяти.

Машины можно обучить автоматизировать научные методы. Формулируя какую-нибудь новую гипотезу, символический «учитель» инвертирует процесс дедукции: я знаю, что Сократ — человек; какая еще информация мне нужна, для того чтобы утверждать, что он смертен? Допустим, нам достаточно знать, что человек смертен; тогда эту гипотезу можно проверить, посмотрев, смертны ли все другие люди из поколения Сократа. Ева, биоробот, созданный в Манчестерском университете в Англии, использовала этот подход для создания потенциального противомаларийного средства. Исходя из данных о конкретном заболевании и знакомства с основами молекулярной биологии, она выдвинула ряд гипотез по поводу того, какие семейства лекарственных средств могли бы работать,

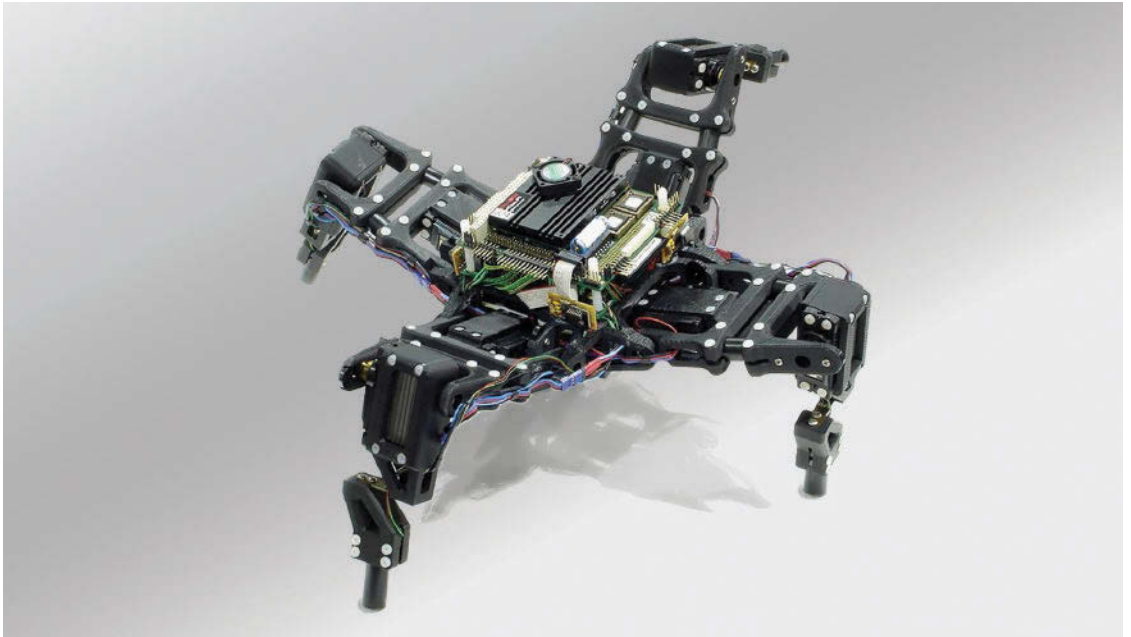
спланировала эксперимент для их проверки, провела его в роболаборатории, многократно проверила и отвергла одну возможность за другой и остановилась, только когда ее все устроило.

Наконец, обучаемые машины могут работать, полагаясь только на математику и прежде всего — на теорию вероятностей Байеса. Согласно этой теории, гипотезам, основанным на нашей информации, приписывается некая исходная вероятность, а затем та, которая лучше других согласуется с имеющимися данными, считается более приемлемой, чем все оставшиеся. Таким образом, каждой из гипотез придается определенный вес — тем больший, чем она вероятнее. Машины, сконструированные по принципу Байеса, иногда ставят бо-

Обучаемые машины могут работать, полагаясь только на математику и прежде всего — на теорию вероятностей Байеса. Сконструированные по принципу Байеса, они иногда ставят более точный диагноз, чем врачи

лее точный диагноз, чем врачи. На их основе создаются фильтры, не пропускающие спам, и системы, которые Google используют для отбора недостающей информации.

У каждого типа обучающих машин есть свои сильные и слабые стороны. Так, глубокое обучение оптимально для решения проблем восприятия, например зрения и слуха, и для распознавания речи, но не для проблем когнитивного характера, таких как приобретение обычных знаний или умение обосновывать те или иные положения. При символическом обучении все



Умный бот. Эта «морская звезда» использует эволюционные алгоритмы для моделирования самой себя. Подобные алгоритмы — один из видов машинного обучения, который в сочетании с другими позволяет создать «универсальный алгоритм», мощнейший инструмент — помощник человека.

наоборот. Эволюционные алгоритмы способны решать более сложные задачи, чем построение нейронных сетей, но для достижения цели требуют очень много времени. Аналоговые методы могут использоваться при обучении совсем небольшое число установок, но становятся в тупик, когда получают о каждой из них слишком много информации. Обучение по Байесу особенно полезно, когда вы имеете небольшой объем данных, и может быть необоснованно дорогим при работе с массивами данных.

Такое разнообразие сфер применения подталкивает разработчиков к заимствованию всего лучшего, на что способны обучаемые машины, составлению некоей компиляции. Подобно инженеру, пытающемуся изготовить ключ, который открывал бы все замки, мы стараемся разработать многоцелевой алгоритм — такой, который мог бы обучить всему, что можно извлечь из имеющихся данных, всей заключенной в них информации.

Нам нужно решить такую же задачу, как и физикам: квантовая механика успешно справляется с описанием природы вещей в микромире, а общая теория относительности — с явлениями космического масштаба, но эти две разномасштабные сущности необходимо как-то примирить. И, следуя Джеймсу Клерку Максвеллу, объединившему свет, электричество и магнетизм и создавшему

теорию электромагнитного поля, разные группы ученых, в том числе и наша из Вашингтонского университета, ищут пути к унификации двух или более парадигм обучения машин. Прогресс в науке развивается нелинейно, и мы не можем предвидеть, когда произойдет полная унификация алгоритмов. Это вовсе не обязательно приведет к появлению новой, доминирующей когорты машин. Вероятнее всего, это ускорит прогресс человечества.

Смена парадигм

Как только мы создадим универсальный алгоритм с использованием огромного массива данных, содержащихся в каждом из нас, системы искусственного интеллекта в принципе смогут построить очень точные и детальные модели любого человека: описать его вкусовые предпочтения и привычки, слабые и сильные стороны; способность к запоминанию, желания, убеждения и личностные характеристики, близких ему людей и животных, его поведение в той или иной ситуации. Подобные модели смогут предугадывать, какой выбор мы сделаем в определенных обстоятельствах.

Есть опасение, что машины с такими возможностями используют новоприобретенные знания для того, чтобы выполнять за нас всю работу, подчинить нас

себе и даже уничтожить. Однако это вряд ли случится, поскольку они не обладают собственной волей. Практически все ИИ-алгоритмы руководствуются тем, что мы в них заложили; они, например, без труда найдут кратчайший путь от отеля до аэропорта, если перед ними поставить такую цель. Отличие алгоритмов высокого уровня от обычных заключается в их изощренности в решении задачи, в поиске оптимальных путей, обычные же алгоритмы просто выполняют определенную, заданную человеком последовательность действий. Даже если они приобретают в ходе работы какие-то новые навыки, цель их действий остается прежней. Решения, которые не направлены к достижению поставленной цели, автоматически отбрасываются. Кроме того, мы всегда можем проверить, соответствует ли продукт, «выпущенный» машиной, нашим желаниям. Мы следим за тем, чтобы машина не вышла в своих действиях за рамки наложенных нами ограничений, что-то вроде соблюдения правил дорожного движения.

Однако когда мы представляем себе искусственный интеллект, мы склонны проецировать на него человеческие качества, такие как воля и сознание. Большинство из нас также больше знакомы с человекоподобными умными машинами, такими как домашние роботы, чем с множеством машин других типов, которые выполняют свою работу «за кулисами». Голливуд усугубляет это восприятие, изображая роботов похожими на людей, — понятная тактика, которая делает историю более привлекательной. Искусственный интеллект просто-напросто позволяет решать сложные задачи — те, для которых не нужна свобода воли. То, что он способен обернуться против нас, не более вероятно, чем если бы наша рука ударила нас. Как и все технологии, он лишь расширяет наши собственные возможности. И чем совершеннее он будет, тем лучше для нас.

Что же нас ждет в будущем? Умные машины освободят нас от многих видов деятельности, но результат будет сходным с тем, к которому привела промышленная революция. 200 лет назад большинство жителей Америки занимались сельским хозяйством. Сегодня почти всех их вытеснили машины, и это не привело к массовой безработице. Скептики возражают, что сейчас другие времена, поскольку ИИ-машины

замещают наши мозги, а не просто мускулы, по сути, не оставляя человеку ничего. Но до тех времен, когда ИИ сможет взять на себя абсолютно все, очень далеко, а скорее всего этого вообще не произойдет. Что касается обозримого будущего, то у ИИ и человека будут разные сферы деятельности. Самое главное, в чем обучаемые машины превзойдут человека, состоит в снижении ценности интеллекта в том смысле, что они опосредуют появление множества экономически полезных применений интеллекта, создавая новые виды деятельности и преобразуя старые так, чтобы они становились более результативными при тех же затратах человеческих ресурсов.

Описанный здесь сценарий можно назвать «сингулярным»; его поклонником и пропагандистом выступает футуролог Рэй Курцвейл (Ray Kurzweil). Это один из самых «быстрых» среди когда-либо существовавших ускорителей технологического прогресса: машины обучают создавать

Ваш цифровой двойник заменит вас практически во всех видах виртуальных взаимодействий: просмотрит все варианты покупки и расскажет вам о них, заменит вас на интервью с работодателем, проанализирует возможные способы лечения и порекомендует наиболее эффективные

еще лучшие машины, те, в свою очередь, — еще более умные, и т.д. Но, как известно, этот процесс не может длиться вечно, поскольку законы физики налагают строгие ограничения на мощность даже квантового компьютера, и похоже, мы не так уж далеки от этого предела. Прогресс в области искусственного интеллекта, как и в любой другой области, в конце концов выходит на плато.

Другая популярная среди футурологов концепция заключается в том, что компьютерные модели будут настолько совершенны, что их нельзя будет отличить от исходного объекта (или субъекта). Согласно этому сценарию, мы можем загрузить себя в определенное облако и постоянно находиться там как часть программы,

не связанные никакими ограничениями физического мира. Одна из проблем этого сценария заключается в том, что он может быть биологически несостоятельным. Чтобы загрузить себя, необходимо вначале с высокой точностью смоделировать каждый нейрон вместе с той информацией, которую он передает. Эта «копия» должна быть настолько реалистична, чтобы предсказания модели нельзя было тут же отделить от поведения настоящих нейронов, — нелегкая задача! Но даже если бы эта опция была достижима, согласились бы вы быть загруженными в программу, будь у вас такой шанс? Где гарантия, что в вашей модели не будет упущена какая-нибудь существенная часть — или что она вообще будет что-либо осознавать? Что если вирус украдет вашу идентичность в полном смысле этого слова? Мне кажется, любой из нас предпочтет остаться в своей «мягкой», углеродной ипостаси — как в шутку это называют компьютерщики, *wetware* (в отличие от *software*).

Учитывая, как быстро совершенствуются обучаемые машины, а прогнозирующие полицейские системы уже используются, можно сказать, что сценарий фильма «Особое мнение» (где людей арестовывают превентивно, когда они как бы задумывают что-то противозаконное), уже не кажется абсурдным

Ищите человека

Искусственный интеллект, в частности обучающие машины, — в буквальном смысле продолжатель эволюции человека. Ричард Докинз в книге «Расширенный фенотип» (*The Extended Phenotype*) показывает, как легко гены животных манипулируют окружающей средой — начиная с кукушечьих яиц и кончая бобровыми плотинами. Любая технология — это расширение человеческого фенотипа, и то, что мы создаем сегодня, — очередной слой нашего технологического экзоскелета. Я думаю, наиболее вероятный сценарий того, как человек

будет использовать искусственный интеллект, далеко выходит за рамки наших сегодняшних спекуляций.

В ближайшее десятилетие у каждого из нас, вероятно, будет свой «цифровой двойник», ИИ-компаньон, гораздо более полезный, чем современный смартфон. Этот двойник не должен будет находиться у вас в кармане или где-нибудь поблизости. Скорее всего он будет «жить» в облаке, так, как это происходит сегодня с большинством ваших данных. Его прообразом может служить виртуальный ассистент, такой как *Siri*, *Alexa* и *Google Assistant*. «Сердцем» двойника будет ваша модель, построенная исходя из всей информации, которую вы поставили когда-либо цифровому миру: от настольного компьютера и различных интернет-сайтов до гаджетов и сенсоров, говорящих устройств и видеокамер.

Чем совершеннее становится алгоритм и чем больше персональных данных в него заложено, тем адекватнее становится наш цифровой двойник. Создав универсальный алгоритм и соединив его с помощью датчиков с вашей чувственно-моторной системой, вы получите цифрового двойника, знающего о вас больше, чем ваши ближайшие друзья.

Сама модель и соответствующая информация будут храниться в «банке данных», сходном с обычным банком, в котором вы держите свои сбережения. Наверняка многие компании хотели бы оказывать подобные услуги. Сооснователь *Google* Сергей Брин заявил, что *Google* желал бы стать «третьей частью вашего мозга», но вы сами вряд ли согласитесь на такое замещение.

Несмотря на все сказанное выше, основная угроза человечеству от искусственного интеллекта связана не с тем, что он самопроизвольно превратится в монстра, а с тем, что человек, который им управляет, воспользуется им ненадлежащим образом. И первейшая задача банка данных будет состоять в том, чтобы вашу модель никто не мог использовать в обход вас. Вместе с банком вы должны тщательно следить за тем, чтобы к модели не получили доступа криминальные элементы. Понадобится создать ИИ-полицию (полицию Тьюринга, как назвал ее Уильям Гибсон в своей книге «Нейромант», вышедшей в 1984 г.), отлавливающую таких мошенников.

Если вы по несчастью живете в авторитарном государстве, то этот сценарий крайне опасен, поскольку власти могут следить за вами и контролировать все ваши

действия. Учитывая, как быстро совершенствуются обучаемые машины, а прогнозирующие полицейские системы уже используются, можно сказать, что сценарий фильма «Особое мнение» (где людей арестовывают превентивно, когда они как бы задумывают что-то противозаконное), уже не кажется абсурдным. Мир начинает жить с цифровыми двойниками прежде, чем все мы освоили хотя бы одного.

Наша первая задача как личностей будет состоять в том, чтобы не подлаживаться к своим цифровым двойникам и не слишком им доверять. Легко забыть, что ИИ подобен погруженному в себя ученому и останется таким и в будущем. Со стороны ИИ кажется беспристрастным, даже совершенным, но он тоже способен ошибаться, как и мы сами, хотя и по-другому. Например, у ИИ отсутствует здравый смысл и он может совершить ошибку, которую мы никогда не сделаем, скажем, принять человека, переходящего улицу, за гонимый ветром пластиковый пакет. Он склонен воспринимать наши инструкции слишком буквально, сообщая нам в точности то, о чем мы спрашивали, вместо того, что мы действительно хотели бы узнать. Поэтому стоит дважды подумать, прежде чем приказать своему беспилотному автомобилю доставить вас в аэропорт вовремя любой ценой.

С житейской точки зрения наш цифровой двойник мог бы заменить нас во всех видах виртуальных взаимодействий. Это не означает, что он вытеснит нас из жизни, но он мог бы принимать за вас решения, если вы заняты до предела или вам не хватает усидчивости или знаний. Он читал бы все книги, доступные на *Amazon*, и предлагал вам те, которые могли бы заинтересовать вас больше всего. Вы хотите купить автомобиль — ваш цифровой двойник просмотрит все варианты и расскажет вам о них. Вы ищете работу — двойник заменит вас на интервью с работодателем и предложит наиболее подходящее место. Допустим, у вас диагностировали рак. Ваш двойник проанализирует возможные способы лечения и порекомендует наиболее эффективные. (Вы сможете также подключить свое цифровое «я» к медицинским исследованиям, если сочтете это этичным.) Если вы ищете партнера для романтических отношений, ваш двойник просмотрит миллионы виртуальных кандидатов. Выбранные им киберпартнеры вполне могут оказаться вашими спутниками в реальной жизни.

По существу, ваш двойник проживает в киберпространстве бесчисленное количество жизней, но та одна, которую проживаете в физическом мире вы, будет лучшей. Будут ли ваши смоделированные жизни хоть сколько-нибудь похожи на «реальную», а ваш кибердвойник обладать самосознанием (как персонажи в некоторых эпизодах сериала «Черное зеркало») — интересный философский вопрос.

Многих беспокоит, что мы отдадим свои жизни под контроль компьютерам. На самом же деле еще большим контролем будем обладать мы, поскольку получим возможность делать выбор, о котором раньше вообще не подозревали. Кроме того, ваша модель будет извлекать уроки из каждого своего виртуального действия, так что со временем научится лучше понимать ваши желания и предлагать варианты, максимально приближенные к тому, чего вы действительно хотите.

Вообще-то мы уже привыкли к тому, что большинство решений мы принимаем без осознанного вмешательства в работу головного мозга. Наш цифровой двойник — это подобие расширенного подсознания, но с одним очень важным отличием: в то время как ваше расширенное подсознание существует изолированно в вашей голове, цифровой двойник непрерывно взаимодействует со своими «коллегами» других людей и других систем. Двойники разных людей будут стремиться обучать друг друга, образуя сообщество моделей, живущих на компьютерных скоростях, бродящих по «саду расходящихся тропок», выясняя, что бы делали мы, оказавшись там. Наши машины будут нашими разведчиками, освещающими путь в будущее для нас как личностей и как вида. Куда они нас приведут? И куда мы сами выберем идти? ■

Перевод: Н.Н. Шафрановская

ДОПОЛНИТЕЛЬНЫЕ ИСТОЧНИКИ

- Квон Д. Самообучающиеся роботы // ВМН, № 5–6, 2018.
- The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Pedro Domingos. Basic Books, 2015.
- The Digital Mind: How Science Is Redefining Humanity. Arlindo Oliveira. MIT Press, 2017.