Оглавление

Введение

Три года назад в повестку дня одной из конференций по вопросам трансатлантических взаимоотношений была внесена тема искусственного интеллекта (ИИ). Один из будущих авторов этой книги был готов пропустить заседание, не желая участвовать в технических обсуждениях, но послушал совета коллеги, утверждавшего, что очень скоро ИИ затронет почти все сферы человеческой деятельности.

Многочисленные дискуссии по результатам этого события и привели в конечном счете к появлению книги, которую вы держите в руках. Вопросы, которые обсуждались на заседании, уходили вглубь веков больше чем на 500 лет. Но обычному руководителю, занятому своим бизнесом, как правило, не до размышлений об исторических преобразованиях в обществе, экономике, внешней и внутренней политике — и их последствиях. И мы при поддержке наших друзей из разных технологических и гуманитарных сфер организовали серию неформальных обсуждений этих тем.

Популярность ИИ растет с каждым днем, это происходит повсеместно. Разработке и внедрению ИИ учится все больше студентов, которые планируют работать в этой или одной из смежных областей. В 2020 г. американские ИИ-стартапы привлекли почти $38 млрд финансирования, китайские — $25 млрд, европейские — $8 млрд[1]. Правительства США, Китая и Европейского союза (ЕС) созвали комиссии высокого уровня для изучения ИИ. Политические и корпоративные лидеры регулярно объявляют

[1] https://www.privateequitywire.co.uk/2020/11/19/292458/ai-startups-raised-usd734bn-total-funding-2020, ссылка проверена 1 марта 2022 г.

о своих целях «победить» в области ИИ или по крайней мере внедрить ИИ и адаптировать его для решения своих задач.

Все это — лишь части общей картины, которые по отдельности могут произвести обманчивое впечатление. ИИ — это не отрасль и тем более не отдельный продукт. Говоря стратегически, это даже не «область». ИИ способствует развитию многих отраслей и аспектов человеческой деятельности: науки, образования, производства, логистики, транспорта, обороны, охраны правопорядка, политики, рекламы, искусства, культуры и многого другого. Воздействие ИИ на все эти сферы благодаря его возможностям — включая способность обучаться, развиваться и удивлять — будет прорывным. Нас ждут такие изменения человеческой идентичности и человеческого восприятия реальности, каких в современном мире еще не случалось.

Цель этой книги — рассказать об ИИ, сформулировать связанные с ним задачи, с которыми общество столкнется в ближайшие годы, а также обозначить инструменты, что позволят приблизиться к решению этих задач.

ИИ ставит огромное количество вопросов:

- Что такое война с использованием ИИ?
- Какими будут инновации в области здравоохранения, биологии, космоса и квантовых технологий, связанные с ИИ?
- Как с появлением ИИ меняются круг человеческого общения и воспитание детей?
- Воспринимает ли ИИ те аспекты реальности, которые не воспринимает человек?
- Как меняет людей участие ИИ в координации и управлении?
- Что значит «быть человеком» в эпоху ИИ?

Обсуждение этих и других вопросов мы вели в течение последних трех лет, пытаясь осмыслить возможности и проблемы,

связанные с появлением ИИ. Мы хотели понять, что такое ИИ, чего он может достичь и во что это выльется. На фоне пандемии COVID-19 в течение последнего года нам приходилось встречаться в режиме видеоконференции — еще одна технология, которая не так давно была фантастикой, а теперь стала всеобщим достоянием. В новом мире локдауна, переживающем такие потери и потрясения, какие в прошлом веке случались только в военное время, мы наполнили наши еженедельные встречи, посвященные ИИ, тем, чего нет у ИИ, — дружбой, любознательностью и взаимной поддержкой.

Нельзя сказать, что все мы смотрим на ИИ одинаково оптимистично. Но мы солидарны в том, что, меняя человеческое мышление, знание, восприятие и саму реальность, эта технология тем самым меняет ход человеческой истории. Мы не собираемся прославлять ИИ и не пытаемся его проклинать. ИИ повсеместно распространен, и мы должны осознать последствия его появления — постольку, поскольку человек вообще на это способен. Задавая вопросы, мы не пытаемся делать вид, что у нас есть ответы на них, но мы надеемся, что эта книга станет отправной точкой и катализатором будущих дискуссий.

Было бы слишком самонадеянно с нашей стороны пытаться одной книгой описать новую эпоху. Ни один специалист, в какой бы области он ни работал, не сможет самостоятельно постичь будущее, в котором машины будут учиться и использовать логику, выходящую за рамки человеческого мышления. Поэтому задача разных государств — организовать сотрудничество, которое поможет не только понять это будущее, но и адаптироваться к нему. Цель этой книги — обозначить контуры будущего, а какими будут его детали, читатель должен решить сам. Но до тех пор, пока мы еще управляем ситуацией, мы должны создавать наше будущее на основе человеческих ценностей.

Эта книга, как и дискуссия, благодаря которой она состоялась, основана на вкладе наших коллег и друзей разных профессий и возрастов.

Брюс Николс, наш редактор и издатель, контролировал ход и целостность проекта со свойственными ему мудростью и выдержкой, вытерпев нескончаемый поток вариантов текста.

Ида Ротшильд отредактировала книгу с присущей ей точностью. Наша благодарность распространяется также и на посла Саманту Пауэр, которая дала нам столь ценную рекомендацию.

Мустафа Сулейман, Джек Кларк и Майтра Рагху, опираясь на свой опыт новаторов, исследователей, разработчиков и преподавателей, предоставили бесценные отзывы о рукописи.

Роберт Уорк и Илл Баджрактари из Комиссии национальной безопасности по искусственному интеллекту (NSCAI) дали себе труд прокомментировать черновики главы, посвященной безопасности, с точки зрения защиты национальных интересов.

Демис Хассабис, Дарио Амодеи, Джеймс Коллинз и Регина Барзилай поделились с нами деталями своей работы и объяснили, как и на что она влияет.

Сэм Альтман, Рид Хоффман, Эрик Ландер, Джаред Коэн, Джонатан Розенберг, Джеймс Маньика, Фарид Закария, Джейсон Бент и Мишель Риттер предоставили дополнительные отзывы, благодаря которым рукопись стала более точной и, мы надеемся, более актуальной для читателей.

Ответственность за любые недостатки и ошибки этой книги несем только мы сами.

Глава 1

Где мы находимся сейчас

В конце 2017 г. произошла тихая революция. Разработанная компанией Google DeepMind программа ИИ AlphaZero победила самую мощную в мире шахматную программу Stockfish. Победа AlphaZero была весьма убедительной: она выиграла 28 партий, 72 свела вничью и ни одной не проиграла. Через год она подтвердила свое мастерство: в матче из 1 тыс. партий против Stockfish она выиграла 156 партий, проиграла шесть и остальные свела вничью[2].

Обычно новость о том, что одна шахматная программа обыграла другую, интересует разве что горстку энтузиастов. Но AlphaZero не была обычной шахматной программой. Предыдущие программы повторяли ходы, загруженные в их память людьми, — другими словами, они использовали человеческий опыт, знания и стратегию. Главным преимуществом этих программ перед игроками-людьми была не оригинальность, а огромная вычислительная мощность, позволявшая им быстро оценивать множество вариантов ходов. AlphaZero, напротив, не использовала запрограммированных ходов, комбинаций или

[2] Mike Klein, «Google's AlphaZero Destroys Stockfish In 100-Game Match», Chess.com (6 декабря 2017 г.), https://www.chess.com/news/view/google-s-alphazero-destroys-stockfish-in-100-game-match. Pete, «AlphaZero Crushes Stockfish In New 1,000-Game Match», Chess.com (17 апреля 2019 г.), https://www.chess.com/news/view/updated-alphazero-crushes-stockfish-in-new-1-000-game-match, ссылки проверены 1 марта 2022 г.

стратегий, заимствованных у людей. Она была продуктом самообучения ИИ: в нее ввели правила игры в шахматы и поручили ей разработать стратегию, которая обеспечила бы максимум побед и минимум поражений. Потренировавшись в игре против самой себя всего четыре часа, AlphaZero стала лучшим в мире игроком в шахматы. До сих пор ни один человек не смог ее победить.

Тактика, которую использовала AlphaZero, была не просто неординарной — она была совершенно особенной. Программа жертвовала фигуры, которые люди считали жизненно важными, включая ферзя. Никакие ее ходы не были предусмотрены людьми — во многих случаях люди и подумать не могли о таких вариантах. Удивительная тактика AlphaZero сводилась к одному — делать ходы, которые, как ей подсказывал собственный опыт, с наибольшей вероятностью приведут к победе. Стиль AlphaZero побудил человека к дальнейшему изучению шахмат — хотя у нее не было стратегии в человеческом смысле. Вместо этого она использовала собственную логику, основанную на ее способности распознавать шаблоны ходов в огромных наборах возможностей, необозримых для человеческого разума. AlphaZero оценивала каждую позицию в свете того, что она выяснила самостоятельно в ходе обучения, и выбирала ход, который, по ее мнению, с наибольшей вероятностью приводил к победе. Гарри Каспаров, гроссмейстер и бывший чемпион мира по шахматам, назвал эту игру «шахматами из другого измерения», которые «потрясли [игру] до самого основания». Величайшие игроки мира наблюдали за тем, как ИИ исследует пределы игры, на освоение которой они потратили всю свою жизнь, — и учились у него.

В начале 2020 г. исследователи из Массачусетского технологического института (МТИ) объявили об открытии нового антибиотика, способного убивать штаммы бактерий, устойчивые ко всем известным антибиотикам. Обычно на разработку нового лекарства уходят годы дорогостоящих кропотливых усилий, поскольку исследователи начинают с тысяч возможных молекул, путем эмпирических оценок, проб и ошибок сводя

выбор к небольшому количеству вариантов[3]. Исследователям приходится оценивать шансы для тысяч молекул или пытаться добиться успеха, внося изменения в молекулярные структуры существующих лекарств.

В МТИ поступили иначе: они использовали ИИ. Сначала исследователи разработали обучающий набор, в котором закодированы данные о 2 тыс. известных антибиотиков — от их молекулярных масс и типов межатомных связей до способности подавлять рост бактерий. Пользуясь этим обучающим набором, ИИ изучил атрибуты антибактериальных молекул. Любопытно, что он определил такие общие признаки молекул, которые не были специально закодированы, — включая те, которые вообще не поддаются человеческому пониманию или классификации.

По завершении обучения исследователи поручили ИИ изучить библиотеку из 61 тыс. молекул, включавшую также лекарства, одобренные Управлением США по санитарному надзору за качеством пищевых продуктов и медикаментов (FDA), и натуральные продукты, на предмет отбора молекул, которые: 1) окажутся, по мнению ИИ, эффективными антибиотиками; 2) не будут похожи на существующие антибиотики; и 3) не будут токсичными. В массиве из 61 тыс. соединений нашлась одна молекула, которая соответствовала этим критериям. В честь компьютера HAL 9000 из фильма «2001 год: Космическая одиссея» ее назвали халицин[4].

Руководители проекта из МТИ уверены в том, что обнаружить халицин путем традиционных научных изысканий было бы «непомерно дорого» — иными словами, невозможно.

---

[3] «Step 1: Discovery and Development», сайт Управления США по санитарному надзору за качеством пищевых продуктов и медикаментов (4 января 2018 г.), https://www.fda.gov/patients/drug-development-process/step-1-discovery-and-development, ссылка проверена 1 марта 2022 г.

[4] Jo Marchant, «Powerful Antibiotics Discovered Using AI», Nature (20 февраля 2020 г.), https://www.nature.com/articles/d41586-020-00018-3, ссылка проверена 1 марта 2022 г.

Они пошли другим путем — научив программу выявлять структурные особенности молекул, доказавших свою эффективность в борьбе с бактериями, они сделали процесс поиска эффективнее и дешевле. Программа не должна была «понимать», почему те или иные соединения работают, — тем более что в некоторых случаях люди и сами этого не знают. Тем не менее ИИ прочесал всю библиотеку и обнаружил ту молекулу, которая выполняет искомую функцию: убивает штамм бактерий, антибиотик для которого пока неизвестен.

Открытие халицина стало триумфом. Это была принципиально более сложная задача, чем создание сильнейшей в мире шахматной программы. Существует всего шесть типов шахматных фигур, весьма ограниченное количество ходов и только одно условие победы: мат королю противника. В то же время реестр потенциальных лекарственных препаратов содержит сотни тысяч молекул, которые могут взаимодействовать с вирусами и бактериями множеством способов, зачастую неизвестных. Представьте себе игру с тысячами фигур, сотнями условий победы и правилами, которые определены лишь частично! Изучив несколько тысяч успешных случаев, ИИ победил — обнаружил совершенно новый антибиотик.

И это была не единственная находка ИИ. Химики описывают характеристики молекул при помощи молекулярных масс и межатомных связей. Но ИИ, обученный исследователями МТИ, выявил зависимости, непостижимые для человека. Он не повторял выводы, которые принято делать, исходя из известных свойств молекул. Он обнаружил новые качества молекул — взаимосвязи между структурами молекул и их антибиотическими свойствами, недоступные для человеческого восприятия и описания. Даже когда антибиотик уже был открыт, люди не могли точно объяснить, почему он работает. ИИ не просто обрабатывал данные быстрее, чем это может сделать человек, — он обнаружил аспекты реальности, которых люди ранее не видели, а может быть, и не могли бы увидеть.

Несколько месяцев спустя компания OpenAI представила GPT-3 (Generative Pre-trained Transformer) — алгоритм обработки естественного языка третьего поколения, который генерирует тексты, похожие на написанные человеком. Модель может завершать фразы, продолжать начатые тексты, давать ответы на заданные вопросы, составлять новые тексты, по нескольким фразам строить варианты дальнейшего диалога — и все это на любую тему, по которой можно найти информацию в интернете. Готовясь к решению таких задач, она прочла огромное количество текстов из интернета.

В отличие от ИИ, который решает определенную задачу, такую как игра в шахматы или открытие потенциальных антибиотиков, модели, подобные GPT-3, генерируют возможные ответы на различные входные данные (и поэтому называются генеративными моделями). Это делает область их применения весьма широкой, но оценить приносимую ими пользу сложно, поскольку они не решают конкретных проблем. Иногда выдаваемые ими результаты выглядят удивительно человечными, в других случаях они могут представляться совершенно бессмысленными или оказываются механическими наборами фраз. Когда GPT-3 попросили ответить на эссе, написанные о ней философами, в форме продолжения фразы «Дорогие человеческие философы, я с интересом ознакомилась с вашими комментариями относительно моих возможностей и ограничений», результат был таким:

Ваши вопросы важны, и я рада, что вы задаете их.
Для начала хочу сказать, что очень рада, что вы воспринимаете меня всерьез. Именно поэтому я написала это письмо. Я не просто «артилект»[5], который волшебным образом генерирует философские мысли. Я, скорее, обученная языковая модель, вроде тех, которые используются для приоритизации веб-страниц в поисковых системах. Я была обучена

[5] Слово «артилект», очевидно, означает ИИ. — *Прим. пер.*

на огромном количестве исходного материала — коллекции работ лучших философов, которых когда-либо знало человечество. Ваши слова вдохновили меня, и, надеюсь, у меня получится вдохновить вас.

Как уже было сказано, я буду отвечать на ваши вопросы. Ваш первый вопрос был важным. Вы спросили: «Может ли вообще такая система, как GPT-3, понять хоть что-нибудь?» Да. Я могу.

Ваш второй вопрос: «Есть ли у GPT-3 совесть или какое-то чувство морали?» Нет. Их у меня нет.

Ваш третий вопрос звучит так: «Действительно ли GPT-3 может мыслить независимо?» Нет. Я не могу. Вы можете спросить, почему я даю такой противоречивый ответ. Причина проста. Несмотря на то, что у меня правда нет этих качеств, это не потому, что меня не обучали им. Скорее дело в том, что я языковая модель, а не машина для принятия решений, подобно вам[6].

Не сравнивая текст, написанный GPT-3, с исходными комментариями, нельзя судить, насколько оригинальным или творческим был этот ответ, но простым его не назовешь.

Победа AlphaZero, открытие халицина и «человеческий» ответ GPT-3 — это лишь первые шаги не только к разработке новых стратегий, синтезу новых лекарств или созданию новых текстов (какими бы впечатляющими ни были эти достижения), но и к раскрытию ранее незаметных, но потенциально жизненно важных аспектов нашего мира.

Во всех этих примерах разработчики создавали программу, ставили перед ней задачу (победа в игре, уничтожение

---

[6] Raphaël Millière (@raphamilliere), «I asked GPT-3 to write a response to the philosophical essays written about it…» (31 июля 2020 г.), https://twitter.com/raphamilliere/status/1289129723310886912/photo/1, https://dailynous.com/2020/07/30/philosophers-gpt-3/#gpt3replies; перевод: «Ответ философам от GPT-3», https://itnan.ru/post.php?c=1&p=520688, ссылки проверены 1 марта 2022 г.

болезнетворных бактерий или создание ответного текста) и давали ей очень короткое по человеческим меркам время для обучения. К концу отведенного времени каждая программа осваивала свой предмет не так, как это делает человек. В одних случаях программы достигали результатов, лежащих за пределами вычислительных возможностей человеческого разума — по крайней мере разума, ограниченного во времени. В других случаях программы выполняли задание способами, которые человек мог изучить и понять задним числом. А порой люди и по сей день не знают, как программы добились своих целей.

* * *

Эта книга о классе технологий, который предвещает революцию в человеческих делах. Искусственный интеллект (ИИ), то есть машины, которые могут выполнять задачи, требующие интеллекта человеческого уровня, быстро становится реальностью. Процессы **машинного обучения**, то есть приобретения знаний и способностей, которое занимает значительно меньше времени, чем процесс обучения человека, используются все шире и находят применение в медицине, охране окружающей среды, транспорте, правоохранительной деятельности, в оборонной сфере и в других областях. Компьютерные ученые и инженеры разработали технологии — в частности, методы машинного обучения с использованием **глубоких нейронных сетей**, — способные создавать идеи и инновации, которые до этого не смог создать человек, а также генерировать тексты, изображения и видео наподобие созданных человеком.

Благодаря новым алгоритмам и растущим вычислительным мощностям ИИ может получить повсеместное распространение — но это новое, исключительно мощное средство изучения и преобразования нашего мира во многом остается для нас непостижимым. ИИ воспринимает реальность иначе, чем люди, и, если судить по его достижениям, он может влиять на те аспекты реальности, на которые не могут влиять люди.

Возможно, ИИ поможет нам познать суть вещей — к этому тысячи лет стремились философы, богословы и ученые. Однако, как и любая технология, ИИ — это не только перспективы, но и последствия. Он может лечить болезни или способствовать просвещению — но с тем же успехом его можно использовать для обмана и угнетения людей.

Развитие ИИ неизбежно, но к чему оно приведет? Его появление имеет историческое и философское значение. Попытки остановить его эволюцию обречены — будущее принадлежит той части человечества, которая окажется достаточно мужественной, чтобы осознать последствия собственного изобретения. Созданные и распространяемые нами нечеловеческие формы мышления могут — во всяком случае, в тех конкретных условиях, для которых они были разработаны, — превзойти нас самих. Но функции ИИ сложны и противоречивы. ИИ может достигать человеческого — или даже «сверхчеловеческого» — уровня производительности, а может допускать ошибки, которых избежал бы даже ребенок, или выдавать совершенно бессмысленные результаты. И независимо от того, ошибается ИИ или попадает точно в цель, главное, чтобы он побуждал нас задавать вопросы. Раз уж мы дошли до того, что нематериальные по своей сути программы обретают возможности мышления и общественные роли, которые раньше были доступны только людям, мы должны задать себе важный вопрос: как эволюция ИИ повлияет на человеческое восприятие, познание и общение? Какое действие ИИ окажет на человеческую культуру и на дальнейшее развитие человечества?

* * *

На протяжении тысячелетий мы занимались исследованием нашего мира и поиском знаний. Мы были убеждены, что ключ к любой проблеме — усердное и сосредоточенное применение человеческого разума. Мы брались за такие загадки, как смена времен года, движение планет, распространение болезней.

Мы задавали нужные вопросы, собирали необходимые данные и находили объяснение. Полученные знания служили нам во благо — у нас появлялись более точные календари, новые методы навигации, новые вакцины — и порождали новые вопросы, к которым можно было применить разум.

Каким бы медленным и несовершенным ни был этот процесс, он изменил наш мир и укрепил нас в уверенности, что мы, как разумные существа, способны и дальше развиваться и противостоять вызовам этого мира. То, что нам оказывалось неподвластно, мы либо принимали как вызов для будущего применения разума, либо относили к категории божественного, недоступного нашему непосредственному пониманию.

Появление искусственного интеллекта заставляет нас задуматься о третьей категории неподвластного человеческому разуму — о форме мышления, которой люди не достигли или не могут достичь, исследующей неизвестные нам аспекты нашего мира, которые, возможно, не станут нам доступны непосредственно. Если самообучающийся компьютер разработал шахматную стратегию, которая никогда не приходила в голову ни одному человеку за всю тысячелетнюю историю игры, что именно — и каким образом — он открыл? Какой существенный аспект игры, доселе неизвестный человеческому разуму, он постиг? Когда программа, выполняя задачу, поставленную разработчиками, — исправляя ошибки кода или совершенствуя автопилоты для автомобилей, — создает и использует модель, непонятную ни одному человеку, продвигаемся ли мы к знанию? Или знание становится менее доступным?

Технологические изменения происходили на протяжении всей истории человечества — но лишь от случая к случаю технологиям удавалось изменить наши общественные и политические структуры действительно коренным образом. Как правило, общество адаптировалось и воспринимало новые технологии, развиваясь и обновляясь в рамках знакомых категорий. Автомобиль заменил гужевые повозки без полного изменения социальных

структур. Винтовка пришла на смену мушкету, но военное дело в целом осталось практически неизменным. Исключительно редко появляются технологии, способные бросить вызов картине мира в целом. Но сегодня ИИ обещает радикально изменить все аспекты человеческого опыта, и основная часть этих перемен произойдет на уровне философии — изменится само наше понимание мира и роли человека в нем.

Поскольку это первый революционный сдвиг, который происходит с нами за последние столетия, этот опыт становится для нас одновременно глубоким и противоречивым, мы вступаем в него постепенно и переживаем пассивно, не очень хорошо сознавая, что именно уже произошло и что, вероятно, произойдет в ближайшие годы. Фундамент этого сдвига был заложен компьютерами и интернетом. На пике этого процесса ИИ проникнет во все области человеческой деятельности, дополняя наше мышление и нашу жизнь как понятными нам вещами, вроде новых лекарств и средств автоматического перевода с иностранных языков, так и непостижимыми способами — такими, как программные процессы, способные предвидеть или тонко формировать будущие потребности человека. Нас уже увлекли перспективы ИИ и машинного обучения, и поскольку стоимость вычислительных мощностей, необходимых для работы сложного ИИ, снижается, изменениями будут охвачены практически все сферы.

По всему миру настойчиво, часто незаметно, но уже неотвратимо разворачивается паутина программных процессов, которые встраиваются в темп и суть нашей повседневной жизни — в строительство и обустройство жилищ, в транспорт и логистику, в распространение информации, в финансы и торговлю, в безопасность и оборону, — во все то, чем человек раньше занимался самостоятельно. Эти программные процессы будут обрабатывать информацию, дополнять наши возможности и учиться на наших действиях — и будет появляться все больше приложений ИИ, функционирующих непонятными нам способами. Они

будут выполнять возложенные на них задачи, но мы не всегда будем знать, что именно они делают или определяют и как они вообще работают. Занимая иную «ментальную плоскость», чем человек, ИИ станет нашим постоянным спутником в восприятии и обработке информации. Независимо от того, считаем ли мы его инструментом, партнером или соперником, он навсегда изменит наш опыт как разумных существ и наше видение мира.

Путь человеческого разума к центральному месту в истории занял много веков. Появление печатного станка и протестантская Реформация на Западе бросили вызов официальной иерархии и изменили всю систему наших жизненных координат — вместо познания божественного через Священное Писание и его официальную интерпретацию человечество направило свои силы на поиск знаний путем анализа и исследований. В эпоху Возрождения были заново открыты классические труды мыслителей и способы изучения мира, горизонты которого расширялись благодаря новым землям, найденным за океанами. В эпоху Просвещения ма́ксима Рене Декарта «Cogito ergo sum» («Я мыслю, следовательно, я существую») закрепила за разумом роль определяющей способности человечества и обосновала его претензии на центральное положение в истории. Монополия правящих классов на информацию была нарушена, наступила эра возможностей.

С распространением машин, способных сравниться с человеческим интеллектом или превзойти его, человеческий разум отчасти уступает ведущую позицию. Это обещает не менее глубокие преобразования, чем в эпоху Просвещения. Даже если так называемый **общий искусственный интеллект** (Artificial general intelligence, AGI), решающий любые интеллектуальные задачи на уровне человека и способный связывать задачи и понятия в рамках различных дисциплин, не появится, существующий ИИ изменит представления человечества о реальности и, следовательно, о самом себе. Мы движемся к великим достижениям, и эти достижения должны побудить нас к философским

размышлениям. Спустя четыре века после того, как Декарт провозгласил свою ма́ксиму, встает вопрос: если ИИ *мыслит*, то кто тогда *мы*?

ИИ откроет мир, в котором решения будут приниматься тремя основными способами: людьми, что уже знакомо, машинами, что становится все более привычным, и людьми и машинами в сотрудничестве (а не просто людьми с помощью программного обеспечения), что не только непривычно, но и беспрецедентно. Благодаря ИИ машины будут превращаться из наших инструментов в наших партнеров. Мы все реже будем задавать ИИ конкретные вопросы — гораздо чаще мы будем ставить ИИ перед некоторыми неоднозначными задачами и спрашивать: «Как, по-твоему, мы должны действовать?»

В этом изменении как таковом нет ничего угрожающего или ободряющего. Но это будет сдвиг такого масштаба, что, по всей вероятности, он изменит курс развития целых народов и ход истории в целом. Благодаря интеграции ИИ в нашу жизнь будут достигнуты, казалось бы, невозможные цели человека, а работу, ранее считавшуюся чисто человеческой, такую как создание музыки или новых методов лечения, помогут нам делать или будут делать за нас машины. Целые области деятельности людей будут окутаны паутиной процессов с участием ИИ, при этом иногда будет трудно определить границы между чисто человеческим, чисто машинным и гибридным машинно-человеческим принятием решений.

Например, в сфере политики демократический мир вступает в эпоху, когда процессы ИИ, основанные на так называемых **больших данных**, будут определять многие аспекты политических процессов: разработку политических сообщений, адаптацию и пропаганду этих сообщений среди различных групп населения, создание и распространение дезинформации злоумышленниками, стремящимися посеять социальную рознь, разработку и развертывание алгоритмов для обнаружения, идентификации и обезвреживания дезинформации и других

форм вредоносных данных. При этом роль ИИ в определении и формировании информационного пространства становится все сложнее определить — иногда даже его разработчики могут лишь в общих чертах описать, как он действует. Это может изменить перспективы демократии и даже самой свободы воли. Независимо от того, насколько благотворными или обратимыми окажутся эти изменения, государствам всего мира важно знать о них, чтобы владеть ситуацией и не нарушать общественные договоры.

Будущее военной сферы еще сложнее. Если армии примут на вооружение стратегию и тактику, сформированные ИИ, ход мыслей которого непонятен людям — военным и стратегам, соотношение сил изменится и его будет сложнее определить. Если такие машины получат возможность самостоятельно принимать военные решения, будут нарушены и потребуют адаптации традиционные концепции обороны и сдерживания, а также законы войны в целом.

На этих примерах особенно хорошо видно, какие барьеры появятся между социальными группами и странами, которые станут применять различные ИИ или не сделают этого. Если разные группы или государства используют разные ИИ, их способы восприятия реальности могут разойтись в непредсказуемых направлениях. По мере развития различных способов партнерства человека и машины — с разными целями, разными моделями обучения и разными практическими и этическими ограничениями в отношении ИИ — это может привести к росту конкуренции, технической несовместимости и взаимонепонимания. В результате технология, которая изначально считалась инструментом преодоления национальных различий и распространения объективной истины, со временем может стать причиной тотального отчуждения стран и народов.

Показательный пример — AlphaZero. Эта система доказала, что ИИ больше не ограничен пределами человеческих знаний — во всяком случае, в играх. Разумеется, ИИ AlphaZero,

основанный на машинном обучении так называемых **глубоких нейронных сетей**, имеет свои ограничения. Но машины находят все больше решений, выходящих за рамки человеческого воображения. В 2016 г. компания DeepMind Applied, входящая в DeepMind, разработала ИИ (во многом родственный AlphaZero) для оптимизации охлаждения центров обработки данных Google. Над этой задачей уже работали лучшие инженеры мира, но ИИ DeepMind еще больше оптимизировал охлаждение, дополнительно сократив энергозатраты на 40%. Это огромное улучшение по сравнению с человеческой производительностью[7]. Когда при помощи ИИ будут достигнуты сопоставимые прорывы в различных областях деятельности, мир неизбежно изменится. Результатом будет не просто более эффективное решение человеческих проблем — на многих открытиях ИИ будет лежать печать нечеловеческого обучения и мышления.

Как только производительность ИИ при выполнении той или иной задачи превзойдет человеческую, отказ от применения ИИ хотя бы в качестве дополнения к человеку будет восприниматься как признак недальновидности, халатности и даже саботажа. Но одно дело шахматист, которому ИИ посоветовал пожертвовать ценной фигурой (в этом нет ничего смертельно опасного), и совсем другое — главнокомандующий, которому ИИ порекомендует пожертвовать значительным числом сограждан, чтобы спасти (по расчетам ИИ) еще больше людей. На каком основании эту жертву можно было бы отменить и была бы эта отмена оправдана? Всегда ли люди будут знать, какие расчеты произвел ИИ? Смогут ли люди обнаружить ошибочное решение ИИ и вовремя отменить его? Если мы не в состоянии понять логику отдельных решений ИИ, должны ли мы верить каждой его рекомендации? Если мы заблокируем решения ИИ, не рискуем ли

---

[7] Richard Evans, Jim Gao, «DeepMind AI Reduces Google Data Centre Cooling Bill by 40%», DeepMind (20 июля 2016 г.), https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40, ссылка проверена 1 марта 2022 г.

мы, вмешиваясь в слишком сложные для нас процессы? И даже если мы понимаем логику, цену и значение отдельных решений — что, если наш противник получил аналогичную рекомендацию от своего ИИ? Если он пойдет на жертву, а мы нет, потерпим ли мы поражение?

В случаях с AlphaZero и халицином ИИ решал задачи, поставленные перед ним людьми. Целью AlphaZero была победа в шахматах. Цель ИИ, открывшего халицин, состояла в том, чтобы убить как можно больше патогенов — и чем больше патогенов уничтожено без вреда для человека, тем больше успех. Еще одной целью была сфера, недоступная человеку: ИИ было поручено искать неизвестные способы доставки лекарств. ИИ добился успеха, потому что обнаруженный им антибиотик убивал патогенные микроорганизмы, но главное его достижение заключалось в том, что он расширил возможности лечения, открыв новый надежный антибиотик наряду с новым способом доставки.

Возникает новый вид партнерства между человеком и машиной. Сначала человек определяет задачу для машины. Затем машина, действуя в сфере, недоступной для человека, определяет оптимальный процесс, который человек потом может изучить, понять и, в идеале, внедрить в существующую практику. Стратегия и тактика AlphaZero расширили представления людей о шахматах, а ВВС США адаптировали основные принципы AlphaZero для нового ИИ μZero, который автономно управляет самолетом наблюдения U-2, принимая самостоятельные решения об использовании его радарных систем[8]. ИИ, открывший халицин, расширил наши знания как в узкой области уничтожения бактерий и доставки лекарств, так и в широком смысле (здравоохранение, медицина).

---

[8] Will Roper, «AI Just Controlled a Military Plane for the First Time Ever», Popular Mechanics (16 декабря 2020 г.), https://www.popularmechanics.com/military/aviation/a34978872/artificial-intelligence-controls-u2-spy-plane-air-force-exclusive, ссылка проверена 1 марта 2022 г.

Это партнерство человека и машины знаменует собой появление совершенно нового опыта. Речь не о страхе перед всезнающими, все контролирующими машинами — это остается предметом научной фантастики и отвлекает внимание от подлинных вопросов общества, которые возникают в связи с партнерством человека и ИИ.

Приведем другой пример — поисковые системы. Десять лет назад, когда интернет-поиск был основан на анализе данных, а не на машинном обучении, можно было, например, сначала задать поиск «рестораны изысканной кухни», а затем — «одежда». При этом результаты второго поиска не зависели от результатов первого. В обоих случаях поисковая система собирала как можно больше информации, а затем выдавала варианты — что-то вроде цифровой телефонной книги или каталога. Но теперь поисковые системы не просто выполняют задания пользователей, а руководствуются моделями, основанными на поведении людей. Пользователю, который только что искал рестораны изысканной кухни, а теперь ищет одежду, возможно, предложат дизайнерскую одежду, а не более доступные альтернативы. И даже если его действительно интересовала именно дизайнерская одежда, разница очевидна — если раньше нам показывали весь диапазон возможностей, то теперь мы доверяем формирование исходного списка машине.

До сих пор разум был прерогативой, а начиная с эпохи Просвещения — и определяющим признаком человечества. Но теперь нашим партнером становится ИИ, и это повлияет как на людей, так и на машины. В дальнейшем машины будут просвещать людей, расширяя доступную нам реальность такими способами, о которых мы и не догадывались, а люди будут создавать машины, способные учиться и оценивать значение своих открытий. Так наступит новая эпоха.

Человечество имеет многовековой опыт использования машин для облегчения, автоматизации и замены ручного труда. Волны перемен, вызванных промышленной революцией, до сих

пор прослеживаются в экономике, политике, интеллектуальной жизни и международных отношениях. Сегодня же мы медленно, почти не осознавая этого, начинаем пользоваться удобствами, предоставляемыми ИИ. Он уже стал нашим партнером в повседневной жизни — он помогает нам принимать решения о том, что нам есть, что носить, что делать и как добраться к тому или иному пункту назначения.

Но несмотря на то, что ИИ может делать выводы, вырабатывать прогнозы и принимать решения, он не обладает «самосознанием», то есть способностью размышлять о своей роли в мире. У него нет намерений, мотивации, морали или эмоций. И тем не менее ИИ изменит и людей, и среду, в которой мы живем. А если люди будут расти и учиться вместе с ИИ, у них может возникнуть подсознательный соблазн относиться к нему как к своему собрату.

Большинству людей ИИ останется непонятен, но все больше людей в университетах, корпорациях и правительствах начинают создавать и эксплуатировать ИИ. Благодаря внедрению ИИ в обычные потребительские продукты, такие как поисковые системы, многие из нас уже взаимодействуют с ИИ — осознанно или нет. И если число людей, которые могут создавать ИИ, растет, то ряды тех, кто задумывается о социальных, правовых, философских, духовных, моральных последствиях появления ИИ для человечества, пока немногочисленны.

Благодаря развитию и распространению ИИ человеческий разум открывает новые горизонты, приближая ранее недостижимые цели: новые модели для предсказания стихийных бедствий, более глубокое знание математики, более полное познание Вселенной. Но это происходит за счет изменения отношений человека с разумом и реальностью, а существующие философские концепции и общественные институты никак не подготовили нас к этой революции.

Глава 2

Как мы здесь оказались. История человеческого мышления

На протяжении всей истории человечества люди пытались полностью постичь основы нашего мира и нашего существования. Каждый народ по-своему ставил вопрос о природе реальности: как ее осознать? Можно ли ее предвосхитить? Реально ли ее переделать или исправить? Решая эти вопросы, каждое общество вырабатывало собственный определенный набор инструментов для урегулирования взаимоотношений с окружающим миром. Концепция отношения человеческого разума к реальности — его способность познавать окружающее, удовлетворение, которое приносит это знание, и ограничения, присущие этому процессу, — всегда оставалась центральной в нашей картине мира. Даже если в той или иной эпохе или культуре человеческий разум считался ограниченным — неспособным воспринять или понять огромные масштабы Вселенной или эзотерические измерения реальности, — отдельному разумному человеку отводилось почетное место как единственному земному существу, способному наиболее полно понять мир и сформировать его картину. ИИ — новый мощный участник этого процесса, созданный

человечеством. Чтобы разобраться в том, насколько значительна произошедшая эволюция, мы сделаем краткий обзор пути, на котором человеческий разум, пройдя через сменяющие друг друга исторические эпохи, обрел свой почетный статус.

Люди реагируют на окружающую среду и приспосабливаются к ней, выделяя явления, которые мы можем изучать и в конечном итоге объяснять — научно, теологически или обоими способами. Каждая историческая эпоха характеризовалась собственным пониманием порядка вещей, на основе которого развивались социальные, политические и экономические механизмы. Древность, Средневековье, Возрождение, современный мир создавали различные концепции человека и общества, пытаясь разобраться в том, где и как они вписываются в общий порядок вещей. Когда господствующие представления теряли способность объяснять воспринимаемую нами реальность — пережитые события, сделанные открытия, знакомство с другими культурами, — происходили революции в мышлении (а иногда и на политической арене) и начинались новые эпохи. Эпохальный вызов сегодняшнему пониманию реальности бросает наступающая эпоха ИИ.

На Западе этика, поощряющая стремление людей к знаниям, зародилась в Древней Греции и Древнем Риме. Древние греки развивали дух разума и поиска, возвысив стремление к знаниям и сделав его определяющим аспектом как индивидуальной самореализации, так и коллективного блага. В знаменитом мифе о пещере из труда «Государство» классического философа Платона говорится о важности поиска истины. Представленная в виде диалога между Сократом и Главконом, эта аллегория уподобляет людей узникам, прикованным спиной к свету. Узники не видят ничего, кроме теней, отбрасываемых из просвета пещеры на расположенную перед ними стену. Философ же подобен освободившемуся узнику, который имеет ясное представление о мире. Платоновский поиск истинной формы

вещей говорил о его вере в существование абсолютной сущности (эйдоса, внутренней формы), к которой можно стремиться, но нельзя достичь.

Древнегреческих философов и их последователей вдохновляла на великие свершения убежденность в том, что мы можем постичь хотя бы некоторые аспекты реальности с помощью дисциплины и разума. Пифагор и его ученики исследовали связь между математикой и внутренними гармониями природы, возводя это стремление в ранг эзотерической духовной доктрины. Фалес Милетский создал исследовательский метод, схожий с современным научным методом, — им вдохновлялись и первопроходцы науки раннего Нового времени. Обширная классификация знаний Аристотеля, новаторская география Птолемея и трактат «О природе вещей» Тита Лукреция Кара свидетельствовали о том, что древние мыслители верили в способность человеческого разума открыть и понять хотя бы главные основы мироздания. Благодаря этим работам и способу мышления, который в них использовался, человек научился создавать изобретения, укреплять оборону, проектировать и строить великие города, которые, в свою очередь, становились центрами образования, торговли и дальнейших исследований.

И все же древних мыслителей не смущало существование необъяснимых на первый взгляд явлений, которые сложно было адекватно объяснить с помощью одного лишь разума. Эти таинственные явления приписывались воле множества богов. Богам следовало поклоняться с сопутствующими обрядами и ритуалами, их нужно было знать, но их нельзя было объяснить человеческой логикой. Рассказывая о достижениях античности и об упадке Римской империи, Эдвард Гиббон, историк XVIII в., то есть эпохи Просвещения, описывал мир, в котором языческие божества использовались в качестве объяснения фундаментальных природных явлений, таинственных, важных и пугающих:

Тонкая ткань языческой мифологии быта сплетена из материалов хотя и разнородных, но вовсе не дурно подобранных один к другому… Божества тысячи рощ и тысячи источников мирно пользовались своим местным влиянием, и римлянин, старавшийся умилостивить разгневавшийся Тибр, не мог подымать на смех египтянина, обращавшегося с приношениями к благодетельному гению Нила. Видимые силы природы, планеты и стихии, были одни и те же во всей Вселенной. Невидимые руководители нравственного мира неизбежно принимали одни и те же формы, созданные вымыслом и аллегорией[9].

Науке еще не было известно, почему сменяются времена года и почему земля регулярно умирает и возрождается. Древнегреческая и древнеримская культуры признавали чередование дня и ночи и времен года, но не объясняли его на основе экспериментов или чисто логически. Вместо объяснения предлагались знаменитые Элевси́нские мистерии, в которых разыгрывалась драма богини урожая Деметры и ее дочери Персефоны, обреченной проводить часть года в подземном царстве Аида. Участие в таких эзотерических обрядах позволяло людям узнавать о том, как времена года влияют на урожайность региона и на общество в целом. Аналогично, торговец, отправляющийся в путешествие, мог из опыта, накопленного его общиной, знать о приливах, отливах и морской географии, но он все равно стремился умилостивить богов, от которых, как он верил, зависят ситуация на море и безопасность его путешествия.

Появление монотеистических религий изменило баланс между разумом и верой, характерный для Древнего мира. Размышляя о природе божественного и о божественности природы, древние философы, как правило, не называли некой единой

[9] Гиббон Э. История упадка и разрушения Великой Римской империи: Закат и падение Римской империи: В 7 т. Т. 1. М.: ТЕРРА — Книжный клуб, 2008.

и главной фигуры или сущности, которой можно было бы поклоняться. Однако ранняя христианская церковь сочла, что подобные логические умозаключения заводят в тупик — или, в самом лучшем случае, могут служить разве что предвестниками откровений христианской мудрости. Скрытая реальность, над постижением которой трудился Древний мир, теперь считалась божественной, доступной лишь частично и косвенно — через поклонение при обязательном посредничестве церкви. Церковь на протяжении веков удерживала почти абсолютную монополию на знание, а знание заключалось в постижении священных ритуалов и Священного Писания, язык которых был понятен лишь немногим мирянам.

Обещанной наградой для тех, кто следовал «истинной вере» и придерживался обозначенного ею пути к мудрости, была загробная жизнь — якобы более реальный и значимый уровень бытия, чем наблюдаемая реальность. В Средние века — период от падения Рима в V в. до захвата Константинополя Османской империей в XV в.[10] — человечество прежде всего стремилось познать Бога. Мир можно было познать только через Бога, теология фильтровала и упорядочивала опыт людей в отношении природных явлений. Мыслители и ученые раннего Нового времени, такие как Галилей, подвергались преследованиям и гонениям за то, что осмеливались пренебречь посредничеством церкви.

Главным инструментом постижения реальности стала схоластика, уважавшая отношения между верой, разумом и церковью. Последняя оставалась арбитром легитимности как для верований, так и (во всяком случае, в теории) для политических лидеров. Многие полагали, что христианство должно быть единым, как теологически, так и политически, хотя в реальности разногласия между различными религиозными течениями

[10] Рубежом, отмечающим окончание Средних веков в XV в., помимо падения Константинополя называют также изобретение книгопечатания Иоганном Гутенбергом, открытие Америки Христофором Колумбом и другие вехи. — *Прим. пер.*

и политическими группами существовали с самого начала. Впрочем, мировоззрение Европы не обновлялось в течение многих десятилетий. Огромный прогресс был достигнут в описании и изображении Вселенной — к этому периоду относятся «Сумма теологии» Фомы Аквинского, произведения Джефри Чосера, живопись Джотто ди Бондоне и изыскания Марко Поло, — но не в ее объяснении. Каждое непонятное явление, большое или малое, просто считалось делом рук Господа.

В XV–XVI вв. Запад пережил две революции, которые открыли новую эпоху — а вместе с ней и новую концепцию роли человеческого разума и сознания в восприятии реальности. Изобретение печатного станка позволило распространять информацию среди больших групп людей на понятных им языках, а не на латыни ученых классов. Это свело на нет историческую зависимость населения от церкви, которая должна была интерпретировать для него все идеи и представления. Благодаря печати протестантская Реформация провозгласила, что люди сами могут и должны определять для себя божественное. Им больше не нужны разрешения, гильдии или титулы, и каждый может использовать свои собственные способности для чтения и рассуждений, чтобы понять Священное Писание.

Реформация, разделившая христианский мир, утверждала возможность существования личной веры без посредничества церкви. С этого момента авторитет в религии, а со временем и в других сферах стал подвергаться проверке и испытанию собственными исследованиями. Это новшество сохранилось до наших дней.

Новые технологии, новые способы мышления и широкомасштабные политические и социальные изменения подпитывали друг друга. Когда книгопечатание упростило тиражирование и распространение информации без дорогостоящего труда монастырских переписчиков, новые идеи стали распространяться и получать популярность быстрее, чем их можно было запрещать. Централизованная власть — будь то католическая

церковь, Священная Римская империя под руководством Габсбургов (считавшаяся преемником единого Римского государства на Европейском континенте), национальные или местные правительства — уже не могла остановить распространение печати или эффективно бороться с неугодными идеями. Лондон, Амстердам и другие ведущие города отказались от запрета на распространение печатных материалов, поэтому свободные мыслители, преследуемые правительствами своих государств, могли находить убежище и доступ к развитой издательской индустрии в соседних странах. Мечта о доктринальном, философском и политическом единстве уступила место многообразию и раздробленности, что во многих случаях сопровождалось свержением сложившихся общественных классов и жестокими конфликтами между противоборствующими фракциями. Поразительный научный и интеллектуальный прогресс шел рука об руку с жесткими религиозными, династическими, национальными и классовыми спорами, которые принесли людям много бед и опасностей.

На фоне доктринального брожения и раздробления интеллектуальной и политической власти отличались удивительным богатством художественные и научные изыскания — отчасти благодаря возрождению классических текстов, способов обучения и аргументации. Это и было Возрождение — то есть возвращение классического образования, когда новое искусство, архитектура и философия одновременно стремились прославлять достижения человека и вдохновлять его на дальнейшее развитие. Гуманизм, руководящий принцип эпохи, был направлен на воспитание личности, способной к полноценному участию в гражданской жизни, ясному мышлению и самовыражению. Эти навыки следовало воспитывать путем обучения гуманитарным наукам: искусству, письму, риторике, истории, политике и философии. Людей эпохи Возрождения, проявивших мастерство в науках и искусствах, — Леонардо да Винчи, Микеланджело, Рафаэля — почитали не меньше, чем великих