ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Представьте, что на дворе 2045 год и вы отправляетесь за покупками. Ваша первая остановка — Центр конструирования разума. Вы входите и видите большое меню с причудливыми названиями, за каждым из которых кроется то или иное расширение возможностей мозга. «Коллективное сознание» — это чип для мозга, позволяющий воспринимать самые потаенные мысли близких людей. «Сад дзен» — микрочип, позволяющий не хуже мастера дзен входить в состояние глубокой медитации. «Живой калькулятор» обеспечит вам математические способности едва ли не гениального уровня. Выбрали бы вы что-нибудь такое и если да, то что именно? Повышенную внимательность? Музыкальные способности, как у Моцарта? Можно заказать одно улучшение, а можно сразу комплект.

А потом вы идете в магазин андроидов. Пора обновить андроида для ведения домашнего хозяйства. Набор видов искусственного интеллекта (будем называть их искинами) обширен и разнообразен. Одни искины позволяют воспринимать или чувствовать то, что людям недоступно, другие обеспечивают связь с базами данных всего интернета. Вы выбираете то, что нужно вашей семье. Весь день у вас уходит на выбор искинов-помощников и конструирование собственного разума.

В этой книге речь пойдет о будущем разума, о том, как наши представления о себе, о разуме и о собственной природе могут изменить будущее — к лучшему или к худшему. Наш мозг приспособлен для жизни в определенных условиях и имеет значительные анатомические и эволюционные ограничения. Искусственный интеллект (ИИ) открывает простор для расширения его возможностей, предлагая новые материалы и принципы работы, а также новаторские пути развития, причем гораздо более быстрые, чем биологическая эволюция. Я называю это захватывающее новое направление *конструированием разума*. Конструирование разума — это форма креационизма, где в роли творца выступаем мы, люди, а не Бог.

Лично меня перспектива конструирования разума заставляет трезво взглянуть на наши нынешние возможности, ведь, если честно, мы не так уж и развиты. Пришелец в фильме Карла Сагана «Контакт» после первой встречи с человеком говорит: «Вы интересный биологический вид. Интересная смесь. Вы способны на такие чудесные сновидения и на такие ужасные кошмары»[1]. Мы ступили на Луну, обуздали энергию атома, но расизм, алчность и жестокость по-прежнему встречаются на каждом шагу. Наше общественное развитие отстает от наших же технических достижений.

На этом фоне мысль о полном отсутствии у нас философского понимания природы разума может показаться мелочью. Однако непонимание в философии тоже обходится недешево, как ясно

демонстрируют две основные сюжетные линии этой книги.

Одна из них всем хорошо знакома. Она присутствует в жизни всегда, от начала до конца: это ваше сознание. Читая это, вы определенным образом ощущаете себя. Вы испытываете телесные ощущения, видите слова на странице и т.д. Ощущаемая сущность вашей ментальной жизни и есть сознание. Без сознания не было бы боли и страданий, не было бы ни радости, ни жгучего любопытства, ни мук скорби. Переживаний, положительных или отрицательных, попросту не существовало бы.

Именно как существо, обладающее сознанием, вы мечтаете об отпуске, прогулке в лесу или изысканных трапезах. Сознание настолько близко, настолько знакомо каждому, что оно естественным образом воспринимается в основном через призму личного опыта. В конце концов, не обязательно читать учебник по нейробиологии, чтобы понять, как ощущается обладание сознанием. Сознание и есть, по существу, внутреннее ощущение. Именно этот базовый элемент — осознанное переживание — служит, на мой взгляд, признаком обладания разумом.

Суть второй сюжетной линии книги заключается в том, что неспособность продумать философские последствия создания искусственного интеллекта (ИИ) может положить конец дальнейшему процветанию существ, обладающих сознанием. При недостаточной осмотрительности реализация технологии ИИ может выйти нам боком, иначе говоря,

вместо упрощения жизни принести страдания, гибель или эксплуатацию людей.

Некоторые связанные с ИИ угрозы процветанию человечества обсуждаются уже давно. Это и хакеры, способные нарушить работу сетей электроснабжения, и сверхразумные автономные системы вооружения, пришедшие словно из фильма «Терминатор». Я же поднимаю вопросы, привлекающие куда меньше внимания. Однако они не менее значимы. Порочная реализация ИИ, о которой я говорю, имеет две разновидности: (1) непредусмотренные ситуации, связанные с созданием мыслящих машин, и (2) последствия радикального расширения возможностей мозга, подобного тому, что предлагается в гипотетическом Центре конструирования разума. Рассмотрим эти опасности по очереди.

Машины, обладающие сознанием?

Предположим, мы умеем создавать совершенные универсальные искины — такие, что могут легко переходить от одного вида интеллектуальной деятельности к другому и даже соперничать с людьми по способности рассуждать. Будет ли это означать, что мы, по существу, умеем создавать машины, обладающие *сознанием*, — машины, которые являются одновременно личностями и субъектами переживания?

Когда дело доходит до вопроса о том, способны мы создать машинное сознание или нет, следует честно признаться, что никто этого не знает. Ясно

одно: вопрос о том, могут ли искины переживать, станет ключом к тому, насколько мы будем ценить их существование. Сознание есть краеугольный камень нашей морали, поскольку оно занимает центральное место в суждении о том, является ли кто-то или что-то личностью, а не просто автоматом. А если искин — это обладающая сознанием сущность, то принуждение его к услужению смахивает на рабство. В конце концов, как бы вы себя чувствовали в магазине андроидов, где на продажу выставлены существа, обладающие сознанием, — существа, умственные способности которых сравнимы с вашими или даже превосходят их?

Если бы я была директором по разработке ИИ компании Google или Facebook, то в будущих проектах мне не хотелось бы попасть в этическую западню из-за создания по недосмотру системы, обладающей сознанием. Появление системы, которая может себя осознавать, могло бы спровоцировать обвинения в ИИ-рабстве и другие кошмары в сфере связи с общественностью. Такое событие могло бы привести даже к запрету использования ИИ-технологий в определенных областях деятельности.

На мой взгляд, все это способно подтолкнуть ИИ-компании к целенаправленной работе над тем, чтобы избежать создания обладающего сознанием искина для определенных задач. Разумеется, для этого сознание должно быть тем, что можно как включить в систему при разработке, так и исключить из нее. Однако оно может быть неизбежным побочным продуктом построения разумной системы или вообще оказаться невозможным.

Не исключено, что в долгосрочной перспективе ситуация изменится не в пользу человека и проблема будет уже не в том, какой вред мы можем нанести искинам, а в том, как искины могут навредить нам. В самом деле, некоторые считают, что синтетический разум будет следующей после человека стадией эволюции разума на Земле. Тогда мы с вами и то, как мы живем и воспринимаем мир, — это всего лишь промежуточный шаг на пути к ИИ, ступенька на эволюционной лестнице. Не случайно Стивен Хокинг, Ник Бостром, Илон Маск, Макс Тегмарк, Билл Гейтс и не только они поднимают «проблему контроля» — вопрос о том, как человек сможет контролировать собственные ИИ-творения, если искины станут умнее нас[2]. Предположим, мы создадим ИИ с уровнем интеллекта, сравнимым с человеческим. Обладая алгоритмами самосовершенствования и способностью к стремительным вычислениям, такой искин мог бы быстро стать намного умнее человека и превратиться в сверхразум — т.е. обрести интеллектуальное превосходство над нами во всех отношениях. Мы, предположительно, не сможем контролировать такой сверхразум, и он, в принципе, получит возможность стереть нас с лица земли. Это лишь один из вариантов вытеснения органического разума искусственными созданиями; в качестве альтернативы человек может слиться с ИИ в результате последовательного усовершенствования мозга.

Проблема контроля стала темой мировых новостей, чему не в последнюю очередь способствовал

недавний бестселлер Ника Бострома «Искусственный интеллект»[*3]. Однако при этом упускается из виду, что сознание может оказаться центральным моментом и в вопросе о том, как ИИ будет оценивать нас. Используя собственный субъективный опыт как трамплин, сверхразумный искин мог бы и у нас увидеть способность к осознанному переживанию. В конце концов, если мы в определенной мере ценим жизнь животных, то делаем это потому, что ощущаем сродство сознания — так, почти любой из нас испытает отвращение при мысли об убийстве шимпанзе, но без всяких угрызений совести съест апельсин. Вот если сверхразумные машины не будут обладать сознанием, — то ли потому, что это невозможно, то ли потому, что их так спроектировали, — мы можем оказаться в беде.

Эти вопросы необходимо рассматривать в еще более широком, поистине вселенском контексте. Как участница двухгодичного проекта в NASA я выдвинула предположение, что схожие события могут происходить и на других планетах; где-то во Вселенной другие биологические виды, возможно, вытесняются искусственным разумом. Занимаясь поисками внеземной жизни, мы должны иметь в виду, что самые разумные инопланетяне могут оказаться *постбиологическими* сущностями, искинами, развившимися на основе биологических цивилизаций. И если эти искины не способны

---

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

обладать сознанием, то с заменой ими биологического разума Вселенная попросту лишается целых популяций существ, обладающих сознанием.

Если сознание, присущее или не присущее ИИ, действительно так важно, как я утверждаю, то нам лучше заранее знать, может ли оно в принципе быть создано и не создали ли его мы, земляне. В последующих главах я расскажу о том, как определить, существует ли искусственное сознание, и коротко опишу тесты, разработанные мной в Институте перспективных исследований в Принстоне.

А теперь рассмотрим идею слияния людей с ИИ. Представьте, что вы находитесь в Центре конструирования разума. Какие расширения возможностей мозга вы выбрали бы из меню — если вообще стали бы это делать? Вы, вероятно, уже начинаете догадываться, что принятие решений при конструировании разума — дело весьма и весьма непростое.

Можно ли слиться с ИИ?

Я не удивлюсь, если идея расширения возможностей мозга при помощи микрочипов кажется вам совершенно пугающей — как, кстати, и мне. Сейчас, когда я пишу это введение, приложения в моем смартфоне, вполне возможно, отслеживают мое местонахождение, слушают мой голос, фиксируют содержание моих поисковых запросов и продают всю эту информацию рекламщикам. Я вроде бы отключила эти функции, но компании,

создающие приложения, делают этот процесс настолько непрозрачным, что уверенности в успехе нет. Если ИИ-компании даже сейчас не уважают нашу частную жизнь, то можно представить себе, какие потенциальные возможности для злоупотреблений появятся, когда самые потаенные наши мысли будут оцифрованы и записаны на микрочипы, а то и доступны где-нибудь в интернете.

Но предположим, что законодательство, регулирующее ИИ, улучшится и наш мозг можно будет защитить от хакеров и алчности корпораций. Тогда, наверное, у вас появится желание расширить возможности мозга — ведь окружающие, похоже, получают немалую пользу от этих технологий. В конце концов, если слияние с ИИ ведет к рождению сверхразума и радикальному увеличению продолжительности жизни, разве этот путь не лучше обычной альтернативы — неизбежной деградации умственных способностей и тела?

Идея о необходимости слияния человечества с ИИ сегодня буквально витает в воздухе. Ее преподносят, с одной стороны, как средство предотвращения вытеснения людей искинами с привычных рабочих мест, а с другой — как путь к сверхразуму и бессмертию. Например, Илон Маск недавно заметил, что люди могут избежать вытеснения искусственным интеллектом при помощи «соединения биологического и машинного интеллекта»[4]. С этой целью он основал новую компанию Neuralink. Одна из ее первых целей — создать «нейронное кружево», своеобразную паутинку, которая будет вводиться в мозг путем инъекции и обеспечивать

его связь с компьютером. Нейронное кружево* и другие усовершенствования, основанные на технологии ИИ, должны, по идее, сделать возможной беспроводную передачу информации из вашего мозга в цифровые устройства или в облако, где доступны большие вычислительные мощности.

Однако мотивы Маска, похоже, не чисто альтруистические. Он проталкивает на рынок линейку продуктов, призванных решить проблему, которую создает сам ИИ. Возможно, его идеи окажутся полезными, но, чтобы понять, так ли это, нужно абстрагироваться от рекламной шумихи. Правительства, общество и даже сами специалисты в области ИИ должны лучше представить себе, что находится на кону.

Например, если ИИ в принципе несовместим с сознанием, то вы, заменив микрочипом часть мозга, отвечающую за сознание, перестанете сознавать себя и ваша жизнь как существа, обладающего сознанием, закончится. Вы станете тем, что философы называют «зомби» — лишенным сознания подобием своего прежнего «я». Даже если такая замена возможна без зомбирования, радикальное

* Нейронное кружево представляет собой гибкую сетку из проволоки диаметром 5 микрон, покрытой пластиковой изоляцией и содержащей сенсоры электрических потенциалов. Смешанная с гелем сетка должна вводиться в мозг через иглу шприца и расправляться там в соответствии с анатомическими особенностями мозговой ткани. Jia Liu et al. Syringe injectable electronics Nat Nanotechnol. 2015 Jul; 10(7): 629–636. doi: 10.1038/nnano.2015.115. — *Прим. науч. ред.*

усовершенствование мозга все равно будет представлять серьезный риск. После серьезных изменений получившаяся личность, возможно, уже не будет вами. Человек, совершенствующий себя таким образом, может незаметно закончить в ходе этого процесса свое существование.

Мой опыт подсказывает, что многие сторонники радикального усовершенствования мозга не осознают, что усовершенствованное существо, возможно, будет уже другим человеком. Они, как правило, считают, что разум — это что-то вроде компьютерной программы. По их мнению, можно радикально улучшить физические компоненты мозга и при этом сохранить ту же самую программу, прежний разум. Они считают, что разум человека, подобно компьютерному файлу, который мы загружаем и выгружаем, как хотим, можно загрузить в облако. Это технофильский путь к бессмертию — если угодно, новая «загробная жизнь» разума, способного пережить тело. Но, какой бы притягательной ни казалась эта технологическая форма бессмертия, мы с вами увидим, что такое представление о разуме глубоко ошибочно.

Так что если через несколько десятилетий вы зайдете в Центр конструирования разума или посетите магазин андроидов, помните, что предлагаемые там ИИ-новинки могут оказаться непригодными к использованию по глубоким философским причинам. *Покупатель, будь осмотрителен*. Впрочем, вы можете усомниться в правильности моего предположения и в том, что совершенный ИИ будет создан. Есть ли основания полагать, что хоть что-то из описанного выше все же произойдет?

ГЛАВА 1

Век ИИ

Возможно, вы даже не задумываетесь о существовании ИИ, но он уже вокруг вас повсюду. Он помогает вам искать информацию в Google. Он одерживает верх над мировыми чемпионами Jeopardy![*] и го. И он совершенствуется с каждой минутой. Но универсального ИИ, т.е. такого, который способен самостоятельно поддержать разумную беседу, включающую идеи из различных областей знания, и даже превзойти человека в сообразительности, у нас нет. Такого рода ИИ фигурирует в кинофильмах «Она», «Из машины» и прочей фантастике, и может показаться, что там он и останется.

Я, однако, подозреваю, что момент его появления не так уж далек от нас. К развитию ИИ нас подталкивают рыночные механизмы и военно-промышленный комплекс — в настоящее время миллиарды долларов тратятся на разработку боевых роботов, умных помощников по

[*] В России эта телеигра получила название «Своя игра». — *Прим. ред.*

дому и суперкомпьютеров, имитирующих работу человеческого мозга. Так, правительство Японии, предвидя недостаток рабочих рук, развернуло программу, результатом которой должны стать андроиды для ухода за престарелыми.

С учетом нынешних более чем стремительных темпов развития ИИ можно ожидать появления универсального искусственного интеллекта уже в ближайшие десятилетия. Универсальный ИИ — это разум, способный, подобно человеческому, соединять воедино неочевидные знания из разных тематических областей и демонстрировать при этом гибкость и здравый смысл. Уже сейчас предполагается, что в ближайшие десятилетия ИИ сделает ненужными многие человеческие профессии. Согласно недавнему обзору, например, самые цитируемые исследователи искусственного интеллекта считают, что ИИ сможет «освоить большинство человеческих профессий по крайней мере так же хорошо, как средний человек» к 2050 г. с вероятностью 50%, а к 2070 г. — с вероятностью 90%[1].

Я уже говорила, что многие эксперты предупреждают о появлении сверхразумного ИИ: искусственного интеллекта, превосходящего умнейших представителей человечества во всех отношениях, включая логические рассуждения и социальные навыки. Сверхразум будет способен уничтожить нас, утверждают они. В противоположность этому Рэй Курцвейл — футурист, занимающий в настоящее время должность технического директора в компании Google, рисует технологическую утопию, которая принесет с собой конец старению,

болезням, бедности и недостатку ресурсов. Курцвейл говорит даже о потенциальных достоинствах дружбы с персонализированными ИИ-системами вроде программы «Саманта» из фильма «Она».

Сингулярность

Курцвейл и другие трансгуманисты сходятся в том, что мы быстро приближаемся к «технологической сингулярности» — точке, где ИИ намного превзойдет человеческий разум и сможет решать задачи, которые мы прежде были не в состоянии решить, с непредсказуемыми последствиями для цивилизации и человеческой природы.

Понятие сингулярности проистекает из математики и физики, в первую очередь из концепции черной дыры. Черные дыры — это «сингулярные» объекты в пространстве и времени, места, где нормальные физические законы рушатся и не работают. По аналогии, технологическая сингулярность должна привести к неуправляемому техническому развитию и вызвать принципиальные изменения в цивилизации. Правила, по которым человечество жило тысячи лет, внезапно перестанут действовать. Ситуация станет непредсказуемой.

Технологические изменения могут оказаться не настолько стремительными, чтобы породить полномасштабную сингулярность, при которой мир меняется чуть ли не в одночасье. Но это не должно отвлекать нас от более серьезного момента: осознания того, что уже в XXI веке мы, вполне возможно, перестанем быть самыми разумными

существами на планете. В будущем лучшими станут умы искусственные.

Мне кажется, мы уже видим причины, по которым искусственный интеллект превзойдет нас во всем. Даже сейчас микрочипы действуют быстрее, чем нейроны. Сегодня, когда я пишу эту главу, самым быстрым в мире является суперкомпьютер Summit в Окриджской национальной лаборатории в штате Теннесси. Его производительность составляет 200 петафлопс, т.е. 200 миллионов миллиардов операций в секунду. То, что Summit делает в мгновение ока, заняло бы всех людей на Земле, производящих по одной вычислительной операции в минуту в течение всего дня, на 305 суток[2].

Разумеется, скорость — это еще не все. Если за мерило взять не арифметические расчеты, то окажется, что ваш мозг обладает куда большей вычислительной мощностью, чем Summit. Ваш мозг — продукт 3,8 млрд лет эволюции (именно во столько оценивается возраст жизни на нашей планете), и его мощь направлена на распознавание закономерностей, быстрое обучение и другие практические задачи выживания*. Индивидуальные нейроны, возможно, работают медленно, но они организованы в многократно запараллеленную структуру, которая до сих пор оставляет современные системы ИИ далеко позади. Но ИИ обладает почти неограниченной возможностью для

* В эволюционном ряду простейший головной мозг впервые встречается у рыб, которые появились 500 млн лет назад. — *Прим. науч. ред.*

развития. Пройдет, может быть, совсем немного времени, и появится суперкомпьютер, способный сравняться по интеллекту с человеческим мозгом или даже превзойти его. Сделано это будет путем воспроизведения мозга и усовершенствования его алгоритмов или через создание новых алгоритмов, вообще не связанных с работой мозга.

Помимо прочего, любой искин можно будет загружать одновременно во множество устройств, легко дублировать и модифицировать, он может уцелеть в условиях, неблагоприятных для биологической жизни, в том числе во время межзвездных путешествий. Наш мозг, несмотря на свою мощность, ограничен объемом черепа и процессами обмена веществ. ИИ, в отличие от нас, может распространить свое присутствие на весь интернет и даже организовать галактический «компьютроний» — громадный суперкомпьютер, использующий в своих расчетах все вещество какой-нибудь галактики. В долгосрочной перспективе о каком бы то ни было состязании не может быть и речи. ИИ будет намного способнее и долговечнее нас.

Ошибка Джетсонов

Все это вовсе не обязательно означает, что мы, люди, потеряем контроль над ИИ и приговорим себя к вымиранию, как утверждают некоторые. Если мы усилим свой интеллект при помощи ИИ-технологий, нам, возможно, удастся не отстать от искусственного интеллекта. Не забывайте, ИИ позволит создавать не только все

более совершенных роботов и суперкомпьютеры. В фильме «Звездные войны» и мультсериале «Джетсоны» люди окружены сложными искинами, притом что сами обходятся без улучшений и остаются просто людьми. Историк Майкл Бесс называет это «ошибкой Джетсонов»[3]. В реальности ИИ не просто преобразует мир. Он преобразует и нас тоже. Нейронное кружево, искусственный гиппокамп*, имплантируемые в мозг микрочипы для лечения аффективных расстройств — вот лишь некоторые из изменяющих разум технологий, которые уже находятся в разработке. Так что Центр конструирования разума не такая уж невероятная выдумка. Напротив, это вполне правдоподобная экстраполяция нынешних технологических тенденций.

Чем дальше, тем больше человеческий мозг рассматривается как нечто, что может быть взломано подобно компьютеру. В одних только США уже немало проектов, где разрабатываются мозговые импланты для лечения психических заболеваний, двигательных расстройств, инсультов, деменции, аутизма и многого другого[4]. Сегодняшние методы оказания медицинской помощи

* Речь идет о работах американского нейрофизиолога Теодора Бергера, показавшего на крысах, а потом и на человеке, что кодированная стимуляция определенных участков гиппокампа может существенно улучшать его функции, связанные с памятью. Theodore W. Berger et al. A Hippocampal Cognitive Prosthesis: Multi-Input, Multi-Output Nonlinear Modeling and VLSI Implementation. IEEE Trans Neural Syst IEEE Trans Neural Syst Rehabil Eng. 2012 Mar; 20(2): 198–211. doi: 10.1109/TNSRE.2012.2189133. — *Прим. науч. ред.*

с неизбежностью найдут продолжение завтра. В конце концов, люди жаждут стать умнее, эффективнее или просто усилить способность радоваться жизни. С этой целью ИИ-компании, такие как Google, Neuralink и Kernel, пытаются определить, как «скрестить» человека с машиной. В ближайшие десятилетия вы, возможно, станете киборгом.

Трансгуманизм

Исследования в этой области начались недавно, но стоит подчеркнуть, что ее основные идеи появились намного раньше и обрели форму философско-культурного движения, известного как *трансгуманизм*. Джулиан Хаксли пустил в оборот термин «трансгуманизм» в 1957 г., когда написал, что в самом близком будущем «биологический вид человека окажется на пороге нового типа существования, отличающегося от нашего не меньше, чем наше существование отличается от существования синантропа»[5].

Трансгуманизм исходит из того, что наш биологический вид в настоящее время находится на сравнительно ранней ступени и что развивающиеся технологии изменят саму эволюцию человека. В перспективе люди будут совершенно непохожи на свое нынешнее воплощение как в физическом, так и в ментальном отношении и приблизятся к персонажам научно-фантастических рассказов. Они получат значительно более развитый интеллект и почти полное бессмертие,

Конец ознакомительного фрагмента.
Приобрести книгу можно
в интернет-магазине
«Электронный универс»
e-Univers.ru