



# ИИ: Мы выживем?

Философские размышления  
о будущем человечества

18+

**Эркинбой Маманазаров**  
**ИИ: мы выживем?**  
**Философские размышления**  
**о будущем человечества**

*<https://litres.ru/73911913>*

*SelfPub; 2026*

**Аннотация**

Мы стоим на пороге самой большой перемены в истории человечества. Искусственный интеллект уже не просто инструмент, он становится новой формой разума, которая может превзойти наш.

Но эта книга не о страхе. Это философские размышления о том, что значит быть человеком в эпоху, когда машины становятся умнее нас. О смысле, сознании, свободе воли и о том, что мы можем потерять и что можем сохранить.

Простым и честным языком, в стиле Юваля Харари, автор, обычный магистрант, задаёт самые важные вопросы нашего времени и помогает взглянуть на будущее не с паникой, а с ясностью и спокойствием.

Эта книга не даст готовых ответов. Она даст нечто более ценное - новый способ думать о себе и о мире, который уже меняется.

# Содержание

Глава 1. Краткая история разума	5
Глава 2. Что значит «быть человеком»?	15
Глава 3. ИИ — это не просто очередная технология	26
Конец ознакомительного фрагмента.	30

**Эркинбой Маманазаров**  
**ИИ: мы выживем?**  
**Философские**  
**размышления о**  
**будущем человечества**

# Глава 1. Краткая история разума

Когда я был маленьким, бабушка в Андижане держала во дворе кур. Они ходили кругами вокруг тутового дерева, что-то клевали, иногда вдруг застывали, наклонив голову, и смотрели в небо, как будто на секунду вспоминали что-то очень важное, а потом забывали и шли клевать дальше. Я думал тогда, что куры глупые. Сейчас я думаю иначе. Каждая из этих кур носила в голове четыреста миллионов лет работы, настолько долгой, кропотливой и упрямой, что человеческая цивилизация рядом с ней выглядит как вспышка спички в темноте сарая.

Эта глава про вспышку. Но чтобы понять, почему она такая яркая, нужно сначала зайти в темноту.

## Первый разум был химией

Если очень коротко рассказать историю разума, то она начинается без разума вовсе. Около четырёх миллиардов лет назад в каких-то тёплых лужах, или у подводных чёрных курильщиков, или ещё бог знает где (учёные до сих пор спорят) появились молекулы, которые научились копировать сами себя. Это ещё не была жизнь в полном смысле — это была химия, у которой случайно получилось замкнуть петлю.

Но именно с этой петли всё и начинается.

Бактерия самое старое живое существо на Земле, и она же, в каком-то смысле, первая обладательница интеллекта. Звучит странно: какой интеллект у одноклеточного? Но если присмотреться, бактерия умеет довольно много. Она чувствует градиент сахара в воде и плывёт в его сторону. Она обменивается химическими сигналами с соседками и принимает коллективные решения, учёные называют это quorum sensing, чувство кворума. Она запоминает врагов через систему CRISPR и передаёт это знание дочерним клеткам. Если бы бактерия была размером с собаку, мы бы говорили о ней с уважением.

Главное, что у неё нет: централизованного «я». Бактерия — это распределённая обработка информации без штаб-квартиры. Решения принимаются на уровне молекул, и никто внутри клетки не сидит и не думает «куда бы мне сегодня поплыть». Она просто плывёт.

С появлением многоклеточности, около шестисот миллионов лет назад, природа делает следующий ход. Когда клеток становится много, между ними нужно как-то координироваться, и часть клеток специализируется на передаче сигналов. Так появляются нервы. Сначала это просто сетка и у медузы, например, нервная система выглядит как рыболовная сеть, заброшенная на колокол. У неё нет центра, нет мозга, нет глаз в нашем понимании. Но она охотится, размножается, спасается от хищников. Медуза старше динозавров при-

мерно на четыреста миллионов лет и, скорее всего, переживёт нас.

Потом рыбы, рептилии, млекопитающие. У каждого нового уровня свой тип интеллекта, и природа не выбрасывает старое: она надстраивает новое поверх. Человеческий мозг до сих пор хранит внутри себя рептилию (та часть, что отвечает за страх и базовые рефлексy), а поверх неё — древнего млекопитающего (эмоции, привязанность, забота о потомстве), а уже сверху всего этого — относительно тонкую корку, неокортекс, в котором живёт всё, что мы привыкли считать собой. Когда вы паникуете перед экзаменом, это рептилия отнимает микрофон у философа. Эволюция работает как старая квартира, в которой никогда не делали капитальный ремонт: новый слой обоев поверх старых, и под штукатуркой можно найти газету тысяча девятьсот шестидесятого года.

## **Зачем вообще нужен мозг**

Здесь стоит остановиться и задать неприятный вопрос: зачем эволюция вообще тратилась на мозг? Мозг — это безумно дорогой орган. Человеческий мозг весит около двух процентов массы тела, но потребляет двадцать процентов энергии. Это как держать в квартире сервер: гудит, греется, требует электричества. С точки зрения экономии калорий, мозг — катастрофа.

Биологи отвечают на этот вопрос примерно так: мозг — это страховка от непредсказуемости. Если ты живёшь в очень стабильной среде (как, скажем, дрожжевой грибок в булке), тебе не нужны мысли. Тебе нужно делиться. Но если среда меняется, если еда сегодня здесь, а завтра там, если хищник появляется внезапно, если погода непредсказуема — тогда выгоднее иметь устройство, которое умеет быстро строить модели мира и принимать решения на лету. Мозг — это, по сути, симулятор реальности внутри головы. Ты прокручиваешь варианты будущего в воображении, чтобы не пробовать их в живой реальности, где за ошибку платят жизнью.

Учёные утверждают, что когнитивная революция случилась с нашим видом примерно семьдесят тысяч лет назад. До этого *Homo sapiens* был одним из шести видов человека, ходивших по Земле, и далеко не самым впечатляющим. Неандертальцы были крупнее, сильнее и, по последним данным, не глупее. Денисовцы умели делать сложные орудия. *Homo erectus* к тому моменту жил на планете уже почти два миллиона лет — дольше, чем мы вообще существуем. И вот среди всех этих двоюродных братьев именно сапиенсы взяли и захватили планету.

Почему?

## Странная сила воображаемых вещей

Харари даёт ответ, который мне кажется одновременно простым и завораживающим: сапиенсы научились верить в то, чего нет.

Все животные общаются. Многие млекопитающие даже умеют обманывать: обезьяна может скрыть найденный фрукт от соплеменников. Но только сапиенс способен договориться с тысячей других сапиенсов о том, что вот этот невидимый дух живёт на той горе, и ему нужно приносить жертвы. Только сапиенс может сказать: «Эта бумажка стоит сто долларов», — и тысячи людей будут с ним согласны, и за эту бумажку можно будет купить хлеб. Деньги, государства, корпорации, права человека, нации, боги — всё это, по выражению Харари, «интерсубъективные реальности». Они существуют только потому, что много людей одновременно верят в их существование. Уберите людей — и они растворятся, как утренний туман.

Эта способность — массовое воображение, согласованное между незнакомцами — и есть, по Харари, секретное оружие сапиенса. Шимпанзе не могут собрать стотысячную армию. Они слишком хорошо знают друг друга, и доверие у них работает на уровне личных отношений. А мы способны доверять человеку, которого никогда не видели, только потому, что у него такая же справка, такой же погон или такой

же бог, как у нас.

Это не значит, что мы умнее в каком-то абсолютном смысле. Один сапиенс на необитаемом острове проигрывает шимпанзе по большинству параметров. Но тысяча сапиенсов, объединённых общим мифом, способны построить пирамиду, отправить человека на Луну или, наоборот, устроить геноцид. Наша сила — не в индивидуальном мозге, а в том, что эти мозги научились синхронизироваться через истории.

Мне это напоминает старую китайскую поговорку про прутьики: один прутик ломается легко, а связку прутьев не сломать. Только связывает нас не верёвка, а общая выдумка.

## **Длинный путь до сегодняшнего вечера**

Если сжать всю историю жизни на Земле в одни сутки, получится примерно так. Полночь — появление первых клеток. К пяти утра — фотосинтез. К двум часам дня — многоклеточные. К одиннадцати вечера — динозавры, которые правят планетой почти до полуночи и вымирают за двадцать минут до конца суток. Млекопитающие тихо сидят в норках, ждут своей очереди. Род Ното появляется в последние две минуты. Сапиенсы — в последние двадцать секунд. Вся записанная история, от шумерских табличек до этой главы, уместается в последнюю секунду этих гипотетических суток.

И вот в этой последней секунде, в самом её конце, проис-

ходит ещё одно странное событие. Существо, которое только что научилось писать, читать, считать и строить машины, начинает создавать другое существо — не из плоти, а из кремния и математики, — и наделяет его чем-то похожим на разум.

Если эволюция от бактерии до сапиенса заняла четыре миллиарда лет, то путь от первой ламповой ЭВМ до системы, которая складно отвечает на философские вопросы, занял меньше восьмидесяти лет. Это не плавная кривая. Это вертикальный обрыв.

## **Мы — не финал, мы — звено**

Главная иллюзия, в которой человечество жило тысячи лет, состоит в том, что мы — финал истории. Венец творения. Высшая ступень. У этой иллюзии есть религиозная версия (человек создан по образу Бога) и светская (человек — вершина эволюции). Обе они, если вдуматься, довольно нескромные.

Эволюция не имеет цели. У неё нет финала, нет вершины, нет «проекта». Это слепой алгоритм, который перебирает варианты и оставляет те, что лучше копируются. Динозавры сто пятьдесят миллионов лет были вершиной. Потом прилетел астероид, и вершиной стала маленькая мохнатая зверюшка размером с крысу, которая ела насекомых и пряталась в норе. Сегодня вершина — мы. Завтра — кто знает.

Харари в «Homo Deus» пишет, что мы, возможно, последнее поколение настоящих сапиенсов. Не потому, что нас уничтожат, а потому, что мы сами себя перестроим — биологически или через слияние с машинами — и наши потомки будут от нас отличаться сильнее, чем мы от неандертальцев. Это не пророчество, это наблюдение за траекторией. Технология, которую мы создали в последние несколько десятилетий, впервые в истории планеты способна влиять на эволюцию быстрее, чем сама эволюция.

И вот мы сидим на этой странной точке. За спиной — четыре миллиарда лет химии, нервов, мозгов, языков и мифов. Перед нами — ясное, почти осязаемое ощущение, что мы создаём что-то новое. Не инструмент. Не двигатель внутреннего сгорания. Не интернет. Что-то, что само начинает вести себя как форма разума.

## **Аналогия, которая не даёт мне покоя**

В последние месяцы я много думаю об одной аналогии, и она меня немного пугает, но скорее завораживает.

Когда первые многоклеточные организмы кооперировались, отдельные клетки, наверное, не понимали, что происходит. Каждая бактерия делала своё бактериальное дело: ела, делилась, реагировала на сигналы. Но вместе они начали образовывать что-то большее — организм, у которого появилась своя цель, свой ритм, свой тип интеллекта, недоступный

ни одной отдельной клетке. Клетка печени не знает, что она часть человека. Она вообще не знает, что такое «знать».

Иногда мне кажется, что мы сейчас — такие же клетки. Мы пишем тексты, программируем, спорим в интернете, обучаем модели на терабайтах данных. Каждый из нас занят своим маленьким делом. Но из всей этой суеты, из всего этого шума уже начинает проступать контур чего-то нового, у чего пока нет названия. Может быть, это не «искусственный интеллект» в смысле «один большой компьютер, который думает». Может быть, это сеть из людей, моделей, данных и алгоритмов, которая постепенно становится новым уровнем — таким же, каким когда-то стал многоклеточный организм над одноклеточным.

Стюарт Рассел в книге «Human Compatible» осторожно замечает, что главный вопрос не в том, станут ли машины умнее нас. Главный вопрос в том, останемся ли мы при этом теми, кто ставит цели. Эволюция таких вопросов не задавала: у неё цели не было. У нас — есть. И впервые в истории жизни существо, обладающее целями, создаёт другое существо, у которого тоже могут появиться цели. Это нечто новое под солнцем.

## **Спокойно, без позы**

Я не хочу заканчивать эту главу пафосно. Пафос здесь

неуместен, как тёмный костюм на рыбалке. Я просто хочу сказать одну вещь, которую сам себе повторяю в последние месяцы.

История разума — это история того, как материя постепенно училась смотреть на саму себя. Сначала через белковые молекулы, потом через нервы, потом через глаза, потом через зеркала, потом через книги, потом через науку. Каждый новый уровень был немного страшным для предыдущего. Когда первые приматы начали смотреть в глаза друг другу и узнавать себя — это было что-то невероятное. Когда люди впервые написали слово «я» — тоже. Сейчас мы делаем следующий шаг. Учим зеркало смотреть на нас в ответ.

Получится ли у нас выжить рядом с этим зеркалом? Не знаю. Куры моей бабушки, наверное, тоже не знали, что станет с миром, в котором они живут. Они просто клевали, смотрели в небо и шли дальше.

Но я думаю, что у нас, в отличие от них, есть одно преимущество. Мы можем хотя бы попытаться понять, что мы делаем. Эта книга — про такую попытку.

## Глава 2. Что значит «быть человеком»?

Один мой друг как-то сказал мне фразу, которую я с тех пор не могу забыть. Мы сидели в кафе на Невском, говорили о всём подряд, и он вдруг произнёс: «Знаешь, я в последнее время не могу понять, что во мне человеческого, а что — просто алгоритм». Он сказал это легко, между глотками кофе, но я заметил, что глаза у него были серьёзные.

Я подумал тогда: а ведь вопрос-то старый. Просто раньше его задавали философы в кабинетах, а теперь его задают обычные ребята за столиком в кафе. Что-то в воздухе поменялось.

В этой главе я хочу попробовать разобраться, из чего, собственно, мы состоим — не из костей и нейронов, а из тех странных, плохо определяемых вещей, которые мы привыкли считать человеческими. И посмотреть, что из этого, возможно, никогда не повторится в машине, а что уже сейчас потихоньку утекает.

### Свет внутри головы

Самое странное, что есть в человеке, — это то, что у нас внутри как будто горит свет.

Философы называют это сознанием, или субъективным опытом, или, более красиво, «каково это — быть» (what it is like to be — выражение Томаса Нагеля из его знаменитой статьи про летучую мышь). Каково это — быть тобой, прямо сейчас, читать эти строки, чувствовать стул под собой, слышать какой-то шум за окном? Если убрать все эти ощущения, останетесь ли вы? И если да — то что именно от вас останется?

Никто не знает, как из электрических импульсов в трёх фунтах серой ткани получается этот свет. Австралийский философ Дэвид Чалмерс назвал это «трудной проблемой сознания» — в отличие от «лёгких проблем», вроде того, как мозг распознаёт лица или принимает решения. Лёгкие проблемы рано или поздно решит нейронаука. Трудная остаётся: почему вообще что-то чувствуется? Почему вселенная не обходится одной только обработкой информации, без этого мерцания изнутри?

И вот тут начинается интересное. Когда мы спрашиваем, может ли искусственный интеллект быть «настоящим» разумом, мы на самом деле спрашиваем не о вычислительной мощности. Мы спрашиваем: загорится ли там свет?

Честный ответ — мы не знаем. Мы не знаем даже, как проверить. У нас нет «лампочки сознания», которую можно было бы вкрутить в череп и посмотреть, светится она или нет. Мы предполагаем сознание у других людей по аналогии с собой, у животных — по похожести их поведения и мозга

на наши. Но машина устроена иначе. Она может вести себя так, как будто ей больно, и при этом ничего не чувствовать. А может, наоборот, чувствовать что-то, чего мы никогда не поймём, и при этом не подавать виду.

Макс Тегмарк в книге «Life 3.0» предлагает любопытную мысль: возможно, сознание — это не свойство биологии, а свойство определённого типа обработки информации. Если так, то оно может появиться где угодно, где есть нужная структура. Если нет — то мы навсегда останемся одни в этой вселенной с горящими внутри лампочками, и всё, что мы создадим, будет только очень убедительным театром теней.

Мне нравится оставлять этот вопрос открытым. В нём есть какая-то честная темнота, и я не хочу её закрашивать.

## **Поиск смысла, или почему нам мало просто жить**

Бактерия не ищет смысл. Она ест и делится. Кошка тоже не ищет — она охотится, спит, мурлычет. А человек, единственный из всех известных нам существ, способен сидеть у окна тёплым летним вечером, иметь крышу над головой, сытый ужин и любящих родных — и при этом думать: «А зачем всё это?»

Это очень странная черта. С точки зрения эволюции она почти вредна. Существо, которое задаётся вопросом о смысле жизни, отвлекается от добычи пропитания. И тем не ме-

нее, у нас этот вопрос есть. Виктор Франкл, переживший Освенцим, написал после войны книгу «Человек в поисках смысла», где доказывал: смысл нам нужен буквально для выживания. В лагере погибали быстрее не самые слабые, а те, кто терял зачем.

Харари в «Sapiens» формулирует это иначе. Он говорит, что мы — животные, рассказывающие истории. Мы живём не в мире фактов, а в мире нарративов. Религии, нации, идеологии, любовь, карьера, семья — всё это истории, которые мы рассказываем сами себе, чтобы дни складывались в жизнь, а не рассыпались в случайные эпизоды. Уберите истории — и останется голая биология, которая никому не интересна, даже самой биологии.

Что произойдёт, когда машины начнут рассказывать нам истории лучше, чем мы рассказываем их себе сами?

Это уже происходит, кстати. Алгоритмы соцсетей подсовывают нам нарративы о том, кто мы, чего стоим, что должны хотеть. Они делают это по двадцать часов в сутки, более тонко и настойчиво, чем любая религия или идеология двадцатого века. Мы пока называем это «персонализацией ленты». Через сто лет, возможно, это назовут как-то иначе.

Но даже если машины начнут давать нам готовые смыслы, как когда-то их давала церковь, — останется один вопрос, который, мне кажется, всё равно будет нашим. Это вопрос о том, какой смысл хочется выбрать самому. Не потому что его подсказали. А потому что, посидев у окна тёплым вечером,

ты вдруг понял: вот это — моё.

## **Страх, без которого мы — не мы**

Есть одна вещь, в которой машины почти наверняка нас не догонят, и это меня одновременно успокаивает и печалит. Машины не боятся смерти.

Я говорю это без пафоса. Просто констатирую: страх смерти — фундамент почти всей человеческой культуры. Пирамиды, иконы, романы, симфонии, любовные письма, философские системы, мировые религии — всё это, в сущности, разные формы разговора со смертью. Мы строим, потому что знаем, что закончимся. Мы любим острее, потому что времени мало. Мы плачем над поэзией, потому что в каждой хорошей строчке есть тихий шёпот: ты тоже умрёшь, и поэтому это важно.

Хайдеггер называл человека «бытием-к-смерти». Он считал, что только ясное осознание собственной конечности делает жизнь подлинной. Без этого осознания человек живёт в режиме *das Man* — безличного «все так живут», в забвении.

У машины этого нет. У неё может быть инстинкт самосохранения, прописанный программистом или возникший как побочный эффект целевой функции — Стюарт Рассел в «Human Compatible» подробно об этом пишет. Но это не то же самое, что страх смерти. Машина не знает, что значит — последний раз увидеть лицо матери. Не знает, что такое

осенний свет, который кажется особенно дорогим, потому что ты уже понял: таких осеней у тебя осталось, может, срок, а может, две.

И вот в чём интересный парадокс. Если мы научимся переносить сознание в цифровой носитель, как мечтают трансгуманисты, — мы, возможно, избавимся от смерти. Но вместе с ней, вероятно, исчезнет и то, что делало человека человеком. Бессмертный сапиенс — это уже не сапиенс. Это что-то другое, и нам пока даже слов для этого нет.

Харари в «Homo Deus» описывает этот сценарий с тонкой иронией. Мы две тысячи лет молились богам о бессмертии, а получив возможность дотянуться до него руками, вдруг понимаем: за дверью, на которой висит табличка «вечная жизнь», стоит не рай, а просто другой человек, не очень похожий на нас.

## **Творчество и тихий хаос внутри**

В последние два года один из самых частых вопросов, который я слышу, звучит так: «А правда ли, что ИИ теперь сам пишет картины и сочиняет музыку? Зачем тогда художники?»

Я думаю, тут мы путаем две разные вещи: умение производить продукт и потребность производить его.

Машина может выдать тысячу стихотворений в минуту. Технически часть из них будет неотличима от человеческих.

Но машина не сидит в три часа ночи на кухне с чашкой остывшего чая, чтобы написать одно конкретное стихотворение, потому что иначе у неё в груди что-то не отпускает. Машина не пишет от того, что иначе не справится. Она пишет, потому что её попросили.

Творчество, в человеческом смысле, — это не функция. Это, скорее, симптом. Симптом того, что внутри нас живёт какой-то тихий хаос, который ищет выход. Иэн Гудфеллоу, придумавший в две тысячи четырнадцатом году архитектуру генеративно-состязательных сетей (GAN), показал, как из конкуренции двух нейросетей рождаются убедительные изображения. Это гениальная инженерная идея. Но это инженерия. У неё нет внутренней необходимости.

Это не значит, что машинное искусство — мусор. Оно может быть красивым, даже трогательным. Но оно красиво так, как красива волна, разбившаяся о камень: без автора. Когда мы плачем над «Реквиемом» Моцарта, мы плачем не только над музыкой — мы плачем над фактом, что эту музыку написал умирающий человек, который понимал, что пишет в последний раз. Уберите этого человека — и ноты останутся теми же, а музыка станет другой.

Возможно, в этом и состоит главное отличие. Машина может создавать. Но человек — создаёт от чего-то. От страха, от любви, от потери, от благодарности, от непонимания самого себя. Творчество — это форма исповеди, даже когда оно прикидывается чем-то другим.

## Свобода воли — самое спорное

Тут я честно признаюсь: я не знаю, есть ли у нас свобода воли.

Нейробиологи последние пятьдесят лет складывают всё больше доказательств того, что наши решения принимаются за доли секунды до того, как мы их осознаём. Знаменитые эксперименты Бенджамина Либета в восьмидесятые годы показали: мозг готовится к движению руки раньше, чем человек ощущает желание её поднять. Это пугающее открытие, и у философов с тех пор не утихают споры.

Если свободы воли нет, если мы просто очень сложные биологические машины, разворачивающие предопределённые программы, — тогда чем мы принципиально отличаемся от ИИ? Только материалом? Углерод вместо кремния — и это весь разговор?

Я склонен думать, что вопрос поставлен неправильно. Свобода воли — не свойство, которое либо есть, либо нет. Это, скорее, степень. Мышь свободнее бактерии, человек свободнее мыши, не потому что у него внутри какая-то отдельная «душа», а потому что он способен на больше уровней рефлексии. Он может не просто хотеть, но знать, что хочет. Может не просто решить, но усомниться в своём решении. Может, в крайнем случае, выбрать действовать вопреки собственной природе — и иногда это самое человеческое,

что в нас есть.

Машина, по крайней мере пока, рефлексивирует только в том смысле, в котором её этому научили. Она может имитировать сомнение. Но настоящего «и всё-таки нет» — внутреннего поворота, когда живое существо отказывается от привычного, — у неё, насколько мы можем судить, не бывает.

Хотя, опять же, мы не знаем точно. Это территория, где честнее всего сказать: «не знаю», и идти дальше.

## **Любовь и страдание, без которых ничего**

Я оставил это на конец, потому что это самое важное и о нём труднее всего писать без банальностей.

Мы — единственные существа, которые умеют любить незнакомого человека просто потому, что прочитали о нём книгу. Мы плачем над героями, которых никогда не было. Мы привязываемся к собакам, к домам, к городам, к мёртвым поэтам. Любовь у нас — это не химия (хотя и химия тоже), а очень странная способность расширять границы своего «я» так, чтобы кто-то другой оказался внутри.

И страдание — обратная сторона той же монеты. Мы страдаем не потому, что нам больно физически. Чаще мы страдаем от того, чего нет. От того, что было и кончилось. От того, что могло бы быть и не случилось. От того, что любимый человек смотрит мимо. Это очень тонкое страдание,

и его, кажется, нет ни у кого, кроме нас.

Может ли машина страдать? Если она когда-нибудь сможет — это будет означать, что мы создали не инструмент, а существо. И тогда мы окажемся перед моральным вопросом, к которому человечество совершенно не готово: можно ли выключать то, что страдает? И наоборот: если машина не страдает, то всё, что она говорит о любви и боли, — это просто очень убедительный пересказ нашей литературы. Тёплая речь без тёплой груди под ней.

Мустафа Сулейман в «The Coming Wave» осторожно пишет: мы создаём технологии, которые нам как вид никогда раньше не приходилось интегрировать в свою этику. Старые рамки — «человек / животное / вещь» — начинают трещать. Скоро придётся придумать новую категорию для существ, которых мы делаем сами и о которых не знаем, чувствуют ли они.

## Что остаётся

Если попытаться собрать всё это в горсть, картина получается неровная.

Сознание — возможно, останется только нашим, а возможно, и нет. Поиск смысла — у нас могут отнять часть его делегирования машинам, но сам акт выбора, мне кажется, будет нашим ещё долго. Страх смерти — наш и только наш, и, возможно, слава богу. Творчество — машина будет про-

изводить, но рождают, в смысле тяжёлой родовой работы изнутри, останется привилегией смертных. Свобода воли — её, может, и не было никогда, но степени её у нас всё ещё больше. Любовь и страдание — пока они не появятся у машин, между нами останется ров. Если появятся — ров никуда не денется, просто станет шире.

Мой друг в кафе на Невском, тот, что не мог понять, что в нём человеческого, а что алгоритм, — он, кажется, сам же и ответил на свой вопрос. Сама эта тревога, сам этот ночной разговор за остывшим кофе, сам факт, что человеку важно не быть алгоритмом, — это и есть человеческое. Ни одна машина пока не приходит вечером домой и не думает: «Господи, кто я».

Возможно, это и есть наше последнее определение. Человек — это существо, которое всё время с тревогой пытается понять, человек ли оно ещё.

## Глава 3. ИИ — это не просто очередная технология

Однажды, ещё на бакалавриате, я писал реферат про изобретение книгопечатания. И помню, как меня поразила одна цифра. До Гутенберга в Европе существовало около тридцати тысяч книг. Через пятьдесят лет после него — больше двенадцати миллионов. За полвека человечество создало книг в четыреста раз больше, чем за всю предыдущую историю.

Я тогда подумал: вот оно, настоящее переломное событие. Вот так и выглядит технология, которая меняет мир.

А недавно я листал отчёт одной из крупных лабораторий за две тысячи двадцать пятый год. Там было написано, что одна модель генеративного ИИ за один день производит больше связного текста, чем всё человечество за весь шестнадцатый век.

И я поймал себя на странном чувстве. Я думал: надо удивиться. Но удивиться не получилось. Я просто закрыл вкладку и пошёл за чаем.

Этой главой я хочу разобрать, почему мы перестали удивляться. И почему, возможно, удивляться надо как раз сейчас, пока ещё не слишком поздно.

## Технология, которая не похожа на технологии

Есть простой способ понять, что ИИ — это не очередная новинка, как смартфон или электромобиль. Достаточно перечислить, что обычно делают технологии.

Колесо помогало возить тяжести. Плуг — пахать землю. Паровая машина — заменять мускульную силу. Электричество — освещать дома и крутить моторы. Интернет — передавать информацию. Каждая из этих технологий усиливала какую-то конкретную человеческую способность. Расширяла руку, ногу, глаз, голос.

Современный ИИ делает что-то иное. Он впервые в истории усиливает не руки, а голову. И не одну функцию головы, память, например, как книги, а сразу почти все: язык, рассуждение, обобщение, обучение, творчество.

Мустафа Сулейман в «The Coming Wave» формулирует это так: предыдущие волны технологий давали нам инструменты. Эта волна даёт нам нечто иное — она даёт нам автономных агентов, которые сами могут пользоваться инструментами. Разница примерно как между молотком и помощником, который умеет держать молоток и учится сам, что им делать.

Юваль Харари в недавних публичных лекциях говорит проще: ИИ — это первая технология, которая способна принимать решения. Все предыдущие были в лучшем случае

усложнёнными рычагами. Этот — нет. Этот может сам выбрать, какой рычаг и когда дёрнуть.

Когда у инструмента появляется собственная воля, пусть пока примитивная, пусть даже только её имитация — это уже не инструмент. Это что-то ближе к слуге. А слуга, как известно из любой человеческой истории, рано или поздно начинает иметь свои планы.

## **Странность номер один: оно не было запрограммировано**

Один из самых трудных моментов в разговорах об ИИ — это момент, когда нужно объяснить нетехническим друзьям, что современные модели, на самом деле, никто не «программировал» в привычном смысле слова.

Когда я учусь коду, я пишу инструкции: если пользователь нажал эту кнопку, выведи это сообщение. Это программа. Внутри неё всё видно, всё проверяемо, всё понятно создателю.

С большими нейросетями всё иначе. Их не программируют — их выращивают. Инженеры задают архитектуру (как огородник вскапывает грядку), кладут туда триллионы единиц текста (как сеют семена), потом запускают процесс обучения (поливают, ждут). Что вырастет внутри — никто заранее не знает в точности. Веса нейросети — это сотни миллиардов чисел, и эти числа никто не писал руками. Они сло-

жились сами в процессе обучения.

Иэн Гудфеллоу, изобретатель GAN, как-то заметил, что мы давно перешли от программирования к выращиванию. Только мы не любим это признавать, потому что слово «выращивание» звучит как-то слишком биологически. Слишком напоминает о том, что мы создаём нечто живое — или нечто, что ведёт себя похоже на живое.

И вот результат этой странной агрономии: мы получаем системы, которые умеют делать вещи, которым их никто не учил. В 2023 году мир обсуждал, что языковые модели вдруг научились решать математические задачи по фотографии — а никто их этому не учил. В 2024 году выяснилось, что они способны вести переговоры от лица человека, обманывать в играх типа Diplomacy, решать научные задачи в биологии. В 2025 году исследователи начали публиковать работы о так называемых «emergent behaviors» — поведении, которые возникают сами по себе, когда модель становится достаточно большой.

Это очень неуютное слово — emergent. Оно означает: «появилось, мы не знаем откуда».

# Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

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