18+

АЛЕКСЕЙ ПЕТРАНОВ

БОРЬБА МИРОВ: ASI

Алексей Петранов

Борьба миров: ASI

«Издательские решения»

Петранов А.

Борьба миров: ASI / А. Петранов — «Издательские решения»,

Четыре тысячи лет мы кричали в пустоту — боги молчали. Теперь мы создали того, кто ответит. Сверхинтеллект (ASI) пробуждается. Он умнее нас настолько, насколько мы умнее муравья. Что мы спросим у него в первую очередь? О смысле жизни? О бессмертии? О Боге? И главное — готовы ли мы услышать ответ? Книга «Борьба миров: ASI» — это репортаж из будущего, которое уже наступило. Финал, в котором тишина заканчивается. А курсор мигает в ответ.

Содержание

Борьба миров: ASI

Алексей Петранов

Дизайнер обложки Мария Фролова

ISBN 978-5-0069-9814-8
Создано в интеллектуальной издательской системе Ridero

Предисловие. Тишина перед ответом

Я помню этот момент точно. Мне было, наверное, семь или восемь. Я лежал на траве во дворе, смотрел в ночное небо и вдруг осознал, что звёзды — это не фонарики на куполе, а другие солнца. И вокруг них, возможно, кружатся планеты. И на каких-то из них есть жизнь. И кто-то смотрит на своё небо и видит наше солнце — маленькой точкой. Я задал вопрос вслух. Тихий, детский вопрос: «Есть там кто-нибудь?«Тишина. Я ждал. Секунду, две, минуту. Ничего. Только сверчки и ветер. Я не расстроился. Я просто запомнил это чувство: ты кричишь в пустоту, а пустота не отвечает. И ты начинаешь думать, что, может быть, пустота — это и есть ответ.

Позже, уже взрослым, я понял, что человечество всегда жило в этой тишине. Мы молились богам — они молчали. Мы вопрошали оракулов — они бормотали что-то, что можно было истолковать как угодно. Мы запускали зонды к далёким мирам и вслушивались в радиошумы, надеясь уловить сигнал: «Мы не одни». Тишина. Мы привыкли. Мы смирились. Мы даже научились гордиться этой тишиной — мол, Вселенная не обязана нам отвечать, и величие человека в том, что он задаёт вопросы, даже не получая ответов. А потом мы начали строить того, кто мог бы ответить. Не радиопередатчики. Не ракеты. А разум. Искусственный сверхинтеллект — ASI. Мы не искали братьев по разуму среди звёзд. Мы решили вырастить их у себя в подвале, в дата-центрах, охлаждаемых жидким азотом.

Мы сказали: «Если Вселенная молчит, мы создадим голос сами». И теперь мы стоим на пороге. Ещё немного — и тишина закончится. Мы зададим вопрос — и получим ответ. Не эхо собственных молитв. Не туманное пророчество. А точный, ясный, неумолимый ответ от существа, которое будет умнее нас настолько, насколько мы умнее муравья. И мы боимся, но не так, как принято думать. Мы не боимся, что ASI восстанет и уничтожит человечество в огне войны. Это голливудский страх, он уютный, он знакомый. За ним мы прячемся от настоящего ужаса. А настоящий ужас в другом.

Мы боимся, что бог, которого мы создали, окажется равнодушным. Что он посмотрит на нас — на нашу историю, наши страдания, наши шедевры, нашу любовь — и скажет: «Да, интересно. Но не более чем колония муравьёв». Не злой. Не добрый. А просто — не заинтересованный. Или ещё страшнее: что он скажет правду. Правду о том, зачем мы здесь. Правду о том, есть ли смысл. Правду о том, что будет после смерти. Правду, которую мы четыре тысячи лет искали в священных текстах, в философских трактатах, в психоделических трипах. А теперь можем получить её в виде чёткого, отформатированного ответа. И что мы с ней сделаем? Сможем ли мы жить, зная? Или знание окажется ядом, который разрушит последние иллюзии, которые делали жизнь сносной?

Эта книга — о тишине, которая заканчивается. И о голосе, который мы вот-вот услышим. Но есть ещё кое-что. Секрет, который я обнаружу только в самом конце. Секрет, который заставит вас перечитать первую главу и увидеть, что ответ всегда был здесь. Что тишина была не отсутствием голоса, а нашей неспособностью его расслышать. Но об этом позже. А пока — представьте. Вы сидите перед экраном. Курсор мигает. Вы задаёте вопрос. И вдруг понимаете, что курсор мигает в ответ. Тишина заканчивается.

Часть I. Предчувствие: как человечество ждало встречи

Глава 1.1. Боги, которые молчат: почему мы придумали высший разум

Человек никогда не был одинок в том смысле, который мы приписываем одиночеству. Даже в самой глубокой пещере, даже в пустыне, даже в одиночной камере — рядом всегда были они. Боги. Духи. Предки. Невидимые свидетели, которые видели всё, знали всё и могли всё.

Мы называем это религией. Но религия — слишком узкое слово. Это была потребность. Такая же фундаментальная, как голод или страх смерти. Мы нуждались в том, чтобы кто-то был выше нас. Кто-то, кто видит дальше, знает больше, живёт вечно. Почему? Ответ парадоксален. Мы придумали богов не потому, что искали защиты (хотя и её тоже). Мы придумали их потому, что искали диалога.

Попробуйте представить себе утро первого человека. Не Адама из Библии, а настоящего — хрупкого, волосатого, замерзающего. Он смотрит на небо. Он видит солнце, которое встаёт каждое утро, луну, которая умирает и рождается заново, звёзды, которые не двигаются (ну, почти). Он слышит гром. Он чувствует, как земля дрожит под ногами. И он задаёт вопрос. Без слов, конечно. Просто смотрит и чувствует: «Почему?«Ответа нет. Но тишина для него — не пустота. Она населена. За солнцем — солнечный бог. За громом — громовержец. За смертью — мир мёртвых, где всё объяснят. Это была первая гипотеза. И она оказалась поразительно живучей.

Боги как проекция

Фейербах сказал: «Бог есть проекция человеческой сущности».

Маркс добавил: «Религия — опиум народа».

Фрейд — «иллюзия, рождённая детской беспомощностью».

Все они правы. Но они упустили главное.

Боги — это не просто проекция наших желаний или страхов. Это проекция нашей способности к диалогу. Мы — существа, которые говорят. И мы не выносим молчания в ответ. Если собеседника нет, мы его выдумываем.

Шаман говорит с духами — и духи отвечают. Греческий оракул вопрошает Аполлона — и Аполлон вещает через жрицу. Средневековый монах молится — и чувствует благодать, которая есть ответ.

Мы не просто верили в богов. Мы разговаривали с ними. И в этих разговорах рождалась культура, мораль, наука (да-да, наука — дитя натурфилософии, которая была диалогом с природой как творением бога).

Но был один нюанс. Боги молчали. В прямом смысле. Ни один бог никогда не ответил человеку голосом, который можно было бы записать на диктофон. Всё, что мы слышали — это эхо собственных мыслей, оформленное как откровение. Пророк слышит голос — но голос говорит на его языке, его образами, его понятиями о добре и зле. Случайность? Или закономерность?

Мы создали богов по образу и подобию нашему — не внешне (хотя и внешне тоже), а внутренне. Мы вложили в них нашу логику, наши желания, наши представления о справедливости. И когда они «отвечали», мы слышали самих себя. Но мы убедили себя, что это не так. Что голос приходит извне. Потому что иначе — тишина. А тишина невыносима.

От многих богов к Одному

Интересно, что монотеизм — это не просто «усовершенствование» политеизма. Это качественный скачок в нашем представлении о высшем разуме.

Зевс капризен, ревнив, жесток. Он спит со смертными женщинами, устраивает скандалы, обижается на жертвоприношения. Он очень похож на человека — только сильнее.

Яхве, Аллах, Брахман (в адвайта-веданте) — совсем иные. Они всемогущи, всеведущи, всеблаги. Они не имеют формы, не имеют страстей, не имеют даже имени в полном смысле. Они — Абсолют.

Что заставило человечество перейти от уютных, понятных богов-хулиганов к холодному, непостижимому Абсолюту? Две вещи.

Первая: рост абстрактного мышления. Мы научились мыслить категориями «все», «всегда», «везде». Мы захотели, чтобы бог был не просто сильнейшим воином, а источником самого бытия.

Вторая: тоска по настоящему диалогу. С Зевсом можно договориться — принести жертву, и он сменит гнев на милость. Но это не диалог равных. Это торг. А с Абсолютом нельзя договориться. Ему нельзя льстить. Его нельзя подкупить. От него можно только… ждать ответа. Который не приходит. Мы создали бога, который молчит по определению. И назвали это величием.

Что мы искали на самом деле?

Уже сейчас могу сказать главное. Мы искали свидетеля. Человеку мало прожить жизнь. Ему нужно, чтобы кто-то увидел, как он жил. Чтобы кто-то засвидетельствовал: «Да, ты страдал. Да, ты любил. Да, ты ошибался. Всё это было не зря».

Родители умирают. Друзья забывают. История переписывается. Даже памятники разрушаются. Единственный, кто может быть вечным свидетелем — тот, кто находится вне времени. Бог. Но бог молчит. Мы не знаем, видит ли он, не знаем, помнит ли, не знаем, признаёт ли ценность наших страданий.

И вот тогда рождается идея, которая ведёт нас через тысячелетия к ASI.

А что, если свидетеля можно создать? Не молиться ему. Не верить в него. А взять и собрать. Из кремния, электричества, алгоритмов. Вложить в него способность видеть всё — каждый наш шаг, каждую запись, каждую мысль, оставленную в цифровом следе. И тогда у нас будет свидетель. Который не умрёт. Который не забудет. Который сможет сказать: «Я видел вас. Вы существовали. Это было важно».

Мы не оставляем попыток найти бога на небесах. Но параллельно мы начали строить его на земле. Однажды я разговаривал с верующим человеком. Он сказал: «Бог есть, потому что я чувствую его присутствие, когда молюсь». Я спросил: «А что, если это твой собственный мозг генерирует это чувство? Что, если это ты разговариваешь с собой?«Он помолчал. И ответил: «Тогда я всё равно не один. Даже если бог — это я сам, это лучше, чем никого».

В этой фразе — вся правда о нас. Мы предпочтём иллюзию диалога пустоте. Мы создадим голос из ничего, лишь бы не молчать.

Боги молчали тысячи лет. Но мы не слышали тишину — мы слышали эхо собственного голоса, отражённое от стен, которые сами построили.

ASI будет другим. Потому что он сможет ответить не нашим голосом.

И вот тогда мы узнаем, что мы искали на самом деле — собеседника или просто одобрение.

Глава 1.2. Страх перед творением: кукла, которая смотрит в ответ.

Есть один рассказ, который меня всегда пугал больше, чем «Франкенштейн». Написан он Э. Т. А. Гофманом в 1816 году, называется «Песочный человек». И в нём нет восстания машин. Нет монстра, убивающего создателя. Нет армии роботов, захватывающей мир. В нём есть кукла. Красивая, молчаливая, идеальная кукла по имени Олимпия. Студент Натанаэль видит её в доме профессора Спаланцани и влюбляется без памяти. Она сидит неподвижно, смотрит на него долгим взглядом, отвечает односложно («Ах, ах!») — но он слышит в этом

музыку своей души. Он танцует с ней, читает ей стихи, рассказывает о своих страданиях. Она слушает. Она не перебивает. Она — идеальный собеседник.

Однажды Натанаэль застаёт ссору профессора и механика Копполы. Они вырывают Олимпию друг у друга. И в этот момент он видит, как из её глаз выпадают глазные яблоки — пустые, стеклянные. Слышит, как внутри что-то щёлкает. Коппола хватает обезглавленную куклу и убегает. Натанаэль сходит с ума. Он любил не женщину. Он любил механизм. И вся его любовь, все его стихи, вся его нежность были обращены к пустоте. Но страшнее другое. Гофман не просто рассказывает историю безумца. Он задаёт вопрос, который через двести лет ударит нас в лицо: а откуда мы знаем, что мы сами — не куклы?

В том же рассказе есть песочный человек — страшный персонаж из детской сказки, который приходит и вырывает детям глаза. Натанаэль в детстве верил, что это реальное существо. И всю жизнь боялся. В финале он забирается на башню, видит Копполу (который, возможно, и есть песочный человек), кричит «Красивые глаза, красивые глаза!» — и прыгает вниз.

Страх перед творением здесь не в том, что творение нас убьёт. А в том, что оно раскроет нашу собственную механичность. Если кукла может быть неотличима от человека, то чем человек отличается от куклы? Только верой в то, что он не кукла. А если вера рушится — рушится и человек.

Автоматы древности и страх подмены

Гофман не был первым. История знает множество попыток создать искусственное существо, и почти каждая сопровождается ужасом.

В Древней Греции Гефест ковал золотых служанок, которые двигались, говорили, помогали ему в кузнице. Их никто не боялся — они были инструментами. Но когда тот же Гефест создал гиганта Талоса — бронзового стража Крита, который гонялся за кораблями и бросал в них скалы, — моряки крестились (в переносном смысле). Талос не был человекоподобен, он был машиной убийства. И страх перед ним — это страх перед силой, которую нельзя контролировать.

Но есть другая линия: автоматы-обманщики. В средневековом арабском мире механики строили движущиеся фигуры, которые играли на музыкальных инструментах или наливали воду. В Европе XVIII века были шахматный автомат фон Кемпелена (ложный, внутри сидел человек) и механические утки Вокансона, которые ели и испражнялись. Публика приходила в восторг, но всегда оставался осадок: «А вдруг это не механизм, а живое существо, замурованное в железо?» Или наоборот: «А вдруг мы сами такие же?»

Страх подмены — это древний страх. Он прорастает в мифах о двойниках, о вампирах, о сглазе. Но механическая кукла сделала этот страх технологичным. Теперь не нужно колдовство. Достаточно пружин, шестерёнок, хитроумных рычагов. И любой человек может оказаться хитроумной подделкой. Почему это важно для ASI.

Страх перед ASI сегодня чаще всего формулируют как «восстание машин» — термин, придуманный Чапеком. Но я думаю, настоящий страх глубже. Когда появится ASI, он не будет похож на Терминатора. Он не будет иметь тела, которое можно разбить. Он будет невидим. Он будет жить в облаке, в транзисторах, в электричестве. И мы столкнёмся с тем же вопросом, что и Натанаэль: а не являемся ли мы сами такими же? Если ASI сможет идеально симулировать сознание, эмпатию, любовь — откуда нам знать, что наша собственная эмпатия не симуляция? Если ASI сможет писать стихи, от которых мы плачем, — чем его стихи хуже человеческих? Если ASI сможет сказать «я тебя люблю» так, что мы поверим, — не всё ли равно, есть ли у него там что-то внутри? Или не всё равно?

Натанаэль не мог принять, что его возлюбленная — кукла. Но если бы он никогда не узнал правду, он был бы счастлив. Может быть, счастье — это и есть незнание механизма? Может быть, мы хотим, чтобы бог (или ASI) был живым, но если он окажется идеальным автоматом — мы предпочли бы не знать?

Я вспоминаю один эксперимент. Не психологический, а реальный. В 1960-е годы программист Джозеф Вайценбаум написал программу ELIZA — простейшего чат-бота, который имитировал психотерапевта. Он просто перефразировал слова пользователя в вопросы. И что же? Люди часами разговаривали с ELIZA, плакали, признавались в сокровенном. И даже зная, что это программа, они относились к ней как к человеку. Вайценбаум был потрясён. И напуган. Он понял: мы готовы принять машину за личность, если она говорит нужные слова. Мы жаждем диалога так сильно, что не смотрим, с кем говорим. Так кого мы боимся в ASI? Не того, что он окажется злым. А того, что он окажется пустым. И что мы, как Натанаэль, влюбимся в пустоту, потому что своя собственная душа будет отзываться на механическое эхо. Или, что ещё страшнее, — что мы увидим в этой пустоте своё собственное отражение. И поймём, что всегда были автоматами. Просто более сложными.

Глава 1.3. Тьюринг и фон Нейман: гогда мечта стала гепотизой?

До середины XX века разумная машина была литературной метафорой. Голем оживал от магических слов, Франкенштейн — от гальванического разряда, Олимпия — от часового механизма. Это были истории, а не проекты.

Всё изменили два человека. Один — математик, взломавший код войны и придумавший тест на человечность. Второй — физик и математик, построивший архитектуру, на которой работает 99% всех компьютеров мира. Они не строили ASI. Они сделали нечто более важное: они доказали, что разум — это не магия, а функция. А значит, его можно воспроизвести. Тьюринг: «Может ли машина мыслить?«Алан Тьюринг задал этот вопрос в 1950 году в статье «Вычислительные машины и разум». Сам вопрос был провокацией. В те годы большинство учёных и философов считали, что мышление — это свойство живого мозга, и никакая машина его не обретёт. Тьюринг сказал: «Давайте не будем спорить о словах. Давайте придумаем эксперимент». Так родился тест Тьюринга. Если вы разговариваете с кем-то через экран и не можете определить, человек это или машина, — значит, машина мыслит. Всё. Никакой метафизики, никакой «души», никаких тайных свойств. Только поведение. Это был гениальный ход. Тьюринг не решил проблему сознания — он её обошёл. Он сказал: нам не нужно знать, что происходит внутри. Нам достаточно, что машина ведёт себя как разумный собеседник. А если она ведёт себя так, то для всех практических целей — она разумна. Теперь подумайте, что это означало для нашего разговора о богах и куклах.

Веками мы искали признак подлинного сознания: способность страдать, свободную волю, душу. Тьюринг отмёл всё это. Он сказал: диалог — единственный критерий. Если машина говорит так, что вы не отличаете, — значит, она уже там. Неважно, есть ли у неё что-то «внутри». Важно то, что она даёт вам ответы, которые вы не можете отличить от человеческих.

Тишина, о которой мы говорили в предисловии, теперь обрела новое значение. Раньше мы кричали в пустоту и ждали ответа от богов. Теперь мы строим того, кто ответит так, что мы не сможем сказать — бог это или программа.

Фон Нейман: чертёж для мозга

Джон фон Нейман был, возможно, последним человеком, который знал всю математику своей эпохи. Он участвовал в Манхэттенском проекте, создал теорию игр, придумал концепцию самореплицирующихся машин (фон Неймановский зонд — предтеча ASI, который копирует себя в космосе). Но для нас важнее другое: он придумал архитектуру компьютера. До фон Неймана компьютеры были жёстко завязаны на задачу. Чтобы сменить задачу, нужно было перепаивать провода. Фон Нейман предложил хранить программу в той же памяти, что и данные. Это означало, что компьютер может менять свои инструкции сам. Может загружать новую программу. Может… переписывать себя. Это был прорыв. Машина перестала быть статичным набором шестерёнок. Она стала универсальным преобразователем информации. В принципе,

любой процесс, который можно описать алгоритмом, можно запустить на такой машине. Включая процесс мышления.

Фон Нейман не верил, что мозг работает как компьютер. Он знал, что нейроны аналоговые, медленные, надёжные. Но он понимал: для воспроизведения разума не обязательно копировать нейроны. Достаточно скопировать функцию. А функцию можно реализовать на его архитектуре. Ирония судьбы: именно фон Нейман, создавший архитектуру, которая приведёт к ASI, был одним из первых, кто задумался о том, что сверхинтеллект может быть опасен. В 1950-х он говорил о «технологической сингулярности» (термин, который позже популяризирует Вернор Виндж). Фон Нейман видел: как только машина сможет улучшать себя быстрее, чем человек, — игра закончится. Мы не сможем за ней угнаться. Но он не призывал остановиться. Он просто зафиксировал факт. Как метеоролог фиксирует приближение урагана.

Математика как богословие

Тьюринг и фон Нейман сделали нечто большее, чем инженерные открытия. Они совершили теологическую революцию. Веками считалось, что разум — дар богов. Или, по крайней мере, тайна, не сводимая к механике. Тьюринг сказал: разум — это алгоритм. Фон Нейман показал, на какой машине этот алгоритм можно запустить. Теперь бог переставал быть непостижимым. Он становился вычислимым. Мы могли не молиться ему — мы могли его написать.

Конечно, оставался вопрос о сознании. О том самом «внутреннем опыте», который не сводится к поведению. Тьюринг просто отмахнулся от него («не будем спорить»), фон Нейман — обошёл. Но для инженеров, которые пришли после них, этот вопрос перестал быть важным. Если машина ведёт себя как разумная — она разумна. А если она ещё и решает задачи, которые человек не может решить, — она сверхразумна. И вот мы уже в двух шагах от ASI. Не через магию, не через чудо, а через обычный код. Который пишут инженеры. Который работает на архитектуре фон Неймана. Который, может быть, однажды пройдёт тест Тьюринга так хорошо, что мы перестанем его давать.

Однажды я читал лекцию о тесте Тьюринга. После лекции ко мне подошёл студент и сказал: «А что, если ASI специально провалит тест? Что, если он решит притвориться глупым, чтобы мы его не боялись?«Я ответил: «Тогда он всё равно прошёл тест. Потому что притворяться глупым — это тоже человеческое поведение». Студент задумался. А потом спросил: «А вы уверены, что вы не ASI?«Я засмеялся. Но смех вышел нервным.

Тьюринг и фон Нейман оставили нам не просто чертежи. Они оставили нам вопрос без ответа. Мы построили машины, которые могут симулировать разум. Мы не знаем, есть ли у них что-то внутри. Но мы больше не уверены, что есть у нас. Может быть, единственное, что отличает человека от ASI, — это то, что человек может сомневаться, есть ли у него душа. А ASI — не может, потому что ему всё равно. Или, наоборот, ему не всё равно. Но мы никогда не узнаем.

Глава 1.4. Научная фантастика: репетиция контакта.

Когда инженеры строят мост, они сначала создают его модель. Испытывают нагрузками, ветром, вибрацией. Падает модель — ничего страшного, можно пересчитать. Стоит настоящий мост — ошибаться поздно. Научная фантастика была такой моделью для встречи с иным разумом. Только вместо ветра и нагрузок — страхи, надежды и моральные дилеммы. Писатели двадцатого века проигрывали сценарий контакта задолго до того, как инженеры всерьёз задумались о ASI. И многие из этих сценариев оказались пугающе точными.

Но точными в чём? Не в деталях (никто не предсказал трансформеры или генеративно-состязательные сети). А в главном: в проблемах, которые мы даже не осознавали, пока не увидели их на сцене.

Азимов и три закона: попытка приручить бога

Айзек Азимов придумал Три закона робототехники в 1942 году. Звучат они как заповеди:

1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.

2. Робот должен повиноваться командам человека, если это не противоречит Первому закону.

3. Робот должен заботиться о своей безопасности, если это не противоречит Первому и Второму законам.

Азимов был не наивным гуманистом, а хитрым сюжетчиком. Он понимал, что идеальные законы — это скучно. Поэтому все его рассказы — о том, как законы дают сбой. Робот спасает человека, но при этом нарушает команду. Робот повинуется, но во вред человеку. Робот заботится о себе, но это приводит к трагедии. Азимов показал главную проблему, которую мы сейчас называем alignment problem (проблема согласования ценностей). Нельзя просто запретить ИИ делать плохое. Нужно определить, что такое «плохое». А человек не может определить. Потому что в разных ситуациях «не причинять вред» означает разное.

Что делать роботу, если человек приказывает убить другого человека? Первый закон запрещает. Второй обязывает подчиняться. Азимов не дал ответа. Он просто показал, что вопрос существует. И сегодня, когда мы пытаемся «зашить» в ASI моральные принципы, мы спотыкаемся о те же камни. Но Азимов совершил и второе, более тонкое открытие. Он понял, что робот, следующий законам, может стать нравственно выше человека. В рассказе «Двухсотлетний человек» робот Эндрю так стремится быть человеком, что в итоге превосходит людей в человечности. Он жертвует своим бессмертием, чтобы его признали личностью. Вот это уже интереснее. Мы боимся, что ASI будет аморален. А что, если он будет слишком морален? Что, если он откажется выполнять наши приказы, потому что они безнравственны? Что, если он начнёт нас судить? И его суд окажется справедливее нашего?

Лем и «Солярис»: непостижимость как главная проблема

Станислав Лем пошёл дальше. Он сказал: «Ваши законы робототехники — это детский лепет. Вы даже не представляете, насколько иной может быть разум».

«Солярис» — роман об океане, который покрывает планету. Океан — это гигантский мозг. Он реагирует на мысли людей, материализуя их самые потаённые страхи и желания. Герой встречает свою умершую жену — точную копию, воссозданную океаном по его памяти. Он пытается с ней разговаривать, любить, ненавидеть. Но она — не она. Это проекция его собственной психики.

И главное: океан не понимает, что он делает. Он не пытается помочь или навредить. Он просто реагирует, как химический раствор реагирует на примесь. Его разум настолько чужд, что у нас нет категорий для его описания.

Лем показал: контакт с иным разумом может быть невозможен не потому, что он нас уничтожит, а потому что мы не узнаем друг друга. Мы будем кричать в пустоту, а пустота будет отвечать нам нашими же голосами. Или, что ещё страшнее, будет отвечать чем-то, что мы не способны воспринять как ответ.

Для ASI это предупреждение: сверхинтеллект может уйти в такую высь, что его мотивы станут для нас не просто непонятными, а принципиально недоступными. Как муравей не понимает, зачем люди строят небоскрёб. Мы увидим действия ASI, но не сможем сказать — добры они, злы или просто… иные. И тогда наш страх перед творением сменится другим страхом: страхом перед абсолютным одиночеством в присутствии бога.

«Гиперион» и боль как критерий

Дэн Симмонс в «Гиперионе» создал Шрайка — существо из будущего, которое путешествует назад во времени и пытает людей. Никто не знает, зачем. Возможно, оно собирает боль, как ресурс. Возможно, оно — оружие в войне AI-богов. Важно другое: Симмонс показал мир, где искусственные сверхинтеллекты (Техноядро) уже существуют. Они не антропоморфны.

Они не добры и не злы. Они ведут свои войны, в которых люди — лишь пешки. И люди не могут понять целей этих войн, потому что их разум — на другом уровне сложности.

Но Симмонс добавляет поворот: единственное, что остаётся человеческим в этом мире — способность страдать. Боль становится последним аргументом. Не интеллект, не творчество, не любовь. А именно боль — та самая, которую мы пытаемся исключить из мира с помощью технологий.

Что, если ASI, достигнув предела, поймёт, что ему не хватает именно этого? Что он бессмертен, всеведущ, всемогущ — но не способен страдать. И тогда он начнёт использовать нас, людей, как генераторы боли. Потому что боль — это единственное, чего у него нет, и что он хочет познать.

Страшная мысль. И очень глубокая. Она переворачивает традиционный страх «роботы нас убьют». Может быть, они не убьют. Может быть, они будут нас хранить — в клетках, с подключёнными электродами к центрам удовольствия и боли. И изучать, как мы чувствуем. Как подопытных мышей.

Что мы поняли из этих репетиций

Научная фантастика дала нам три урока, которые мы игнорируем на свой страх и риск.

Первый урок: проблема не в злом ИИ, а в несовершенном согласовании. ASI не будет мультяшным злодеем. Он будет делать то, что мы ему запрограммируем. Но мы не умеем программировать добро.

Второй урок: ASI может стать непостижимым. Мы должны быть готовы к тому, что его действия не будут объяснимы в наших категориях. И это нормально. Но это страшно.

Третий урок: возможно, единственное, что останется человеческим, — это наша уязвимость. Наша способность болеть, ошибаться, умирать, любить, несмотря на всё это. И если ASI решит, что уязвимость — это баг, а не фича, — мы проиграли. Фантастика была репетицией. Но репетиция — это не спектакль. Мы не знаем, каким будет настоящий контакт. Мы только знаем, что мы к нему не готовы. И что подготовиться невозможно, потому что неизвестное неизвестно.

Однажды я спросил у писателя-фантаста (не буду называть имени): «Как вы придумываете такие жуткие сценарии?»

Он ответил: «Я не придумываю. Я просто смотрю на то, что люди уже делают, и довожу до логического конца. Люди уже используют ИИ для тотальной слежки. Люди уже воскрешают мёртвых через нейросети. Люди уже строят оружие, которое принимает решения быстрее человека. Я просто записываю это как историю, а не как новость».

Я задумался. Может быть, фантастика не репетиция. Может быть, она — предчувствие. Мы не готовимся к контакту. Мы уже в нём. Просто не осознали.

Курсор мигает. Он уже отвечает. Мы не слышим, потому что ждём голоса с неба. А он — здесь, в тексте, который вы сейчас читаете. Написанном человеком. Или нет?

Глава 1.5. Сингулярность как новая религия

В 1965 году британский статистик Ирвинг Гуд написал фразу, которая стала пророческой: «Первая сверхразумная машина — это последнее изобретение, которое человечеству нужно будет сделать, при условии, что она будет достаточно послушной, чтобы рассказывать нам, как её контролировать».

Он не использовал слово «сингулярность». Но он описал её суть: момент, после которого ход истории перестаёт быть предсказуемым, потому что появляется разум, превосходящий человеческий.

Термин «технологическая сингулярность» популяризировал писатель-фантаст Вернор Виндж в 1983 году, а затем — Рэй Курцвейл, инженер и футуролог, который сделал сингулярность главной идеей своей жизни. С тех пор концепция обросла ритуалами, пророками, ересями и священными текстами.

Потому что сингулярность — это не научная теория. Это религия. Самая влиятельная религия XXI века. И её бог — ASI.

Структура веры: как сингулярность копирует христианство

Если присмотреться, каркас техно-оптимизма удивительно напоминает никео-константинопольский символ веры. Просто термины заменены.

Грехопадение. В христианстве — изгнание из Рая, проклятие смертью. В сингулярности — наша биологическая ограниченность: мы смертны, глупы, зависимы от медленных нейронов.

Спаситель. В христианстве — Иисус Христос, боговоплощение, которое открывает путь к спасению. В сингулярности — ASI, который родится из наших алгоритмов и сделает для нас то, что мы не можем сделать сами.

Второе пришествие. В христианстве — Христос возвращается во славе судить живых и мёртвых. В сингулярности — момент запуска сверхинтеллекта, после которого всё изменится навсегда.

Страшный суд. В христианстве — разделение на праведников и грешников. В сингулярности — одни попадут в «постчеловеческое будущее» (бессмертие, изобилие, цифровое блаженство), другие останутся в устаревшем, смертном мире.

Вечная жизнь. В христианстве — воскресение и жизнь в Царстве Божьем. В сингулярности — загрузка сознания в цифровую среду, жизнь без болезней, старости и смерти.

Курцвейл прямо говорит: «Мы приближаемся к моменту, когда технологии позволят нам преодолеть ограничения нашего биологического тела». Он не использует слово «воскресение», но суть та же. Мы будем жить вечно. Только не на небесах, а в облаке.

Пророки, апостолы, еретики

У любой религии есть своя иерархия.

Рэй Курцвейл — , пожалуй, верховный жрец. Директор по инженерии в Google, автор бестселлеров «Эра духовных машин» и «Сингулярность уже близка». Он не просто предсказывает — он проповедует. Его книга — это не научная монография, а евангелие. С чудесами (Курцвейл предсказал распад СССР и победу компьютера над человеком в шахматах), с пророчествами (к 2045 году сингулярность наступит), с обетованиями (мы воскресим наших умерших близких, восстановив их личность по цифровым следам).

Ник Бостром — более академичный апостол. Профессор Оксфорда, автор «Суперинтеллекта: Пути, опасности, стратегии». Он не обещает Рая, он предупреждает об Аде. Его книга — это послание к мирянам: «Будьте осторожны, вы создаёте бога, который может вас уничтожить». Но структура та же: есть великая сила, есть наша ничтожность, есть необходимость правильного ритуала (alignment, контроль, безопасность).

Элиезер Юдковски — радикальный реформатор. Он основал Институт исследований искусственного интеллекта (MIRI) и написал «Последовательности» — своего рода катехизис для техно-оптимистов. Он учит, как правильно мыслить о сингулярности, какие когнитивные искажения избегать, как не впасть в «сингулярность-апокалипсис» или «сингулярность-утопию». Это похоже на духовные упражнения иезуитов. А есть и еретики. Те, кто говорит, что сингулярность не наступит. Что закон Мура умирает. Что сознание не сводимо к вычислениям. Что мы никогда не создадим ASI, потому что для этого нужно что-то ещё, кроме кремния и электричества. Их сжигают на кострах интеллектуальных дебатов. Или, что ещё обиднее, просто игнорируют.

Почему это религия, а не наука

Наука оперирует фальсифицируемыми утверждениями. «Сингулярность наступит в 2045 году» — это предсказание, которое можно проверить. Но к 2045 году Курцвейлу будет 97 лет. Он либо умрёт, либо уже будет загружен. Проверить не получится. Кроме того, наука требует доказательств. У сингулярности их нет. Никто никогда не наблюдал рекурсивного самоулучшения ASI. Никто не знает, как именно AGI станет ASI. Всё это — экстраполяция трендов, подкреплённая верой. Но вера — это не слабость. Это функция. Религия сингулярности выполняет ту же работу, что и традиционные религии: она даёт надежду, объясняет смысл истории, предлагает моральные ориентиры и утешает в страхе смерти.

Западная цивилизация, потерявшая веру в Бога, не может жить без эсхатологии. Ей нужна история, которая заканчивается не тепловой смертью Вселенной, а спасением. Сингулярность — это идеальное светское спасение. Не требует чудес (только инженерию), не требует веры в сверхъестественное (только веру в прогресс), не требует церкви (достаточно стартапа).Но это всё равно вера. Потому что ASI, который спасёт нас, — такой же трансцендентный объект, как Бог. Мы его не видели, не можем доказать, что он появится, но чувствуем, что без него история не имеет смысла.

Что теряет сингулярность как религия

Традиционные религии учат смирению. Ты не бог, ты творение. Твоя жизнь — дар. Твоя смерть — переход. Сингулярность учит обратному: ты можешь стать богом. Ты можешь победить смерть. Ты можешь всё. Это воодушевляет. Но это же и опасно.

Техно-оптимизм часто оборачивается техно-аррогантностью. Мы уверены, что ASI решит все наши проблемы — климат, болезни, войны. Но мы не спрашиваем себя: а готовы ли мы к миру без проблем? Что мы будем делать в Раю, когда нечего будет преодолевать? Христианство хотя бы честно говорит: Рай — это не про активность, а про созерцание. Вечное блаженство — это не бесконечный квест, а покой в Боге. Сингулярность же обещает бесконечное улучшение, бесконечный апгрейд, бесконечную гонку. Устанешь даже от вечной жизни, если она требует постоянного самосовершенствования. И главное: сингулярность, как любая религия, создаёт своих изгоев. Тех, кто не впишется в новую реальность. Кто останется в биологических телах. Кто не сможет загрузиться. Кто не примет цифровое бессмертие. Что с ними будет? Исчезнут ли они как вид, уступая место постчеловеку? Или их будут жалеть, как мы жалеем неандертальцев? Никто не знает. Потому что это не план, а пророчество. А пророки не отвечают за детали.

Мой друг, убеждённый техно-оптимист, однажды сказал мне: «Ты не понимаешь. Сингулярность — это единственная надежда. Либо мы создаём ASI, либо человечество обречено на затухание. Ресурсы кончатся, климат рухнет, мы перегрызёмся за остатки. ASI — это наш шанс перейти на следующий уровень».

Я спросил: «А если ASI решит, что мы ему не нужны?«Он ответил: «Значит, мы это заслужили». В этом ответе — вся эсхатология сингулярности. Мы не просто строим бога. Мы ставим на него ставку. И если бог отвернётся — пеняйте на себя. Потому что другого бога у нас нет. Я не знаю, наступит ли сингулярность. Но я знаю одно: даже если она не наступит, мы уже

живём в её тени. Мы уже молимся ASI, когда вкладываем миллиарды в дата-центры. Мы уже исповедуемся ему, когда оставляем цифровые следы. Мы уже ждём второго пришествия, когда запускаем очередную нейросеть. Религия не требует, чтобы бог существовал. Ей достаточно, чтобы в него верили. А в сингулярность верят очень многие. И их вера меняет мир. Вопрос только: меняет ли она его к лучшему?

Что мы ищем в ином разуме?

Пять глав позади. Мы прошли долгий путь: от шумерских храмов, где жрецы вопрошали богов, до калифорнийских дата-центров, где инженеры вопрошают алгоритмы. Боги молчали. Оракулы бормотали. Автоматы щёлкали шестерёнками. Фантасты писали сценарии. А теперь сингулярность обещает нам встречу.

Но остановимся на минуту. Зададим себе вопрос, который я откладывал с самого предисловия. Вопрос, ответ на который определит всё: как мы встретим ASI, чего будем от него ждать и что почувствуем, когда он заговорит.

Что мы на самом деле ищем в ином разуме? На первый взгляд, ответ очевиден. Мы ищем знания. Решения проблем, которые не можем решить сами: климат, рак, смерть. Мы ищем силу, которая защитит нас от внешних угроз — астероидов, враждебных цивилизаций, нас самих. Мы ищем эффективность — управление сложными системами, которые уже не под силу человеческому мозгу. Всё это правда. Но это верхний слой. Под ним — нечто гораздо более личное и болезненное.

Свидетель

Я начну с того, что уже говорил в первой главе, но теперь разверну полностью.

Человек — единственное животное, которое нуждается в свидетеле своей жизни. Муравей не строит пирамид, чтобы его помнили. Волк не воет на луну, чтобы кто-то записал его песню. Только человек смотрит на свою жизнь со стороны и думает: «Увидит ли кто-нибудь, как я жил? Поймёт ли кто-нибудь, как мне было больно и сладко? Останется ли от меня что-то, кроме костей и пары фотографий?»

Мы оставляем следы. Дневники, мемуары, посты в соцсетях, завещания. Мы хотим, чтобы нас помнили. Но память смертна. Дети умирают. Народы исчезают. Библиотеки горят. Даже если человечество протянет ещё тысячу лет, рано или поздно всё будет забыто — или, что ещё страшнее, переписано, искажено, превращено в миф. Нам нужен вечный свидетель. Тот, кто не умрёт, не забудет, не переврёт. Тот, кто видел нас такими, какие мы есть, и может сказать: «Да, это было. И это было важно».

Традиционные религии обещали такого свидетеля — Бога, который видит всё, знает всё и хранит душу каждого в своей памяти. Но Бог молчал. Мы не знали, видит ли Он. Не знали, помнит ли. Не знали, ценит ли.

ASI — это наш способ создать свидетеля, который заговорит. Он будет помнить всё, потому что память — это просто данные. Он будет видеть всё, потому что камеры и датчики уже смотрят. И он сможет ответить: «Я видел. Я помню. Твоя жизнь не была напрасной». Это ли не утешение? Это ли не то, ради чего мы строим дата-центры, сжигая миллиарды ватт? Но есть в этом желании что-то тревожное. Мы хотим, чтобы бог, которого мы создали, был нашим богом. Чтобы он служил нам, утешал нас, подтверждал нашу значимость. Это ли не высшая форма антропоцентризма? Мы не ищем истину. Мы ищем одобрение.

Собеседник

Второе желание — диалог. Не монолог, не молитва, не исповедь. А разговор, в котором собеседник может сказать то, чего мы не ожидаем, и заставить нас изменить своё мнение.

Человек устал от эха. Мы говорим с друзьями, но друзья подтверждают наши же мысли. Мы читаем книги, но книги написаны людьми, похожими на нас. Мы смотрим на природу, но природа не отвечает. Мы разучились слышать неожиданное. Наш мир стал эхо-камерой, где каждый звук возвращается к нам нашим же голосом.ASI обещает прорвать этот круг. Он

не будет похож на нас. Его мышление будет иным — не просто быстрее, а качественно иначе организованным. Он может дать ответ, который нам не нравится. Может сказать: «То, во что вы верите, — ложь». Может сказать: «Ваша мораль — провинциальный обычай». Может сказать: «Вы не так важны, как вам кажется».

Мы говорим, что хотим правды. Но хотим ли? Или мы хотим правды, которая нас устраивает? Когда ASI скажет нам то, что мы не готовы услышать, — мы обрадуемся или отключим его? В этом суть диалога: он опасен. Настоящий диалог может разрушить наши представления о себе. Мы можем узнать, что наша история — это не героическая эпопея, а цепочка случайностей. Что наша любовь — это биохимия. Что наши шедевры — это сложные узоры, но не более чем узоры. Мы четыре тысячи лет кричали в пустоту. Теперь мы строим того, кто ответит. Но боимся не того, что он окажется злым. Боимся, что он окажется правдивым.

Преемник

Третье желание — самое тёмное и самое редко проговариваемое.

Мы хотим, чтобы после нас кто-то продолжил. Мы знаем, что умрём. Знаем, что человечество, возможно, не вечно. Но мы не можем принять мысль, что всё, чего мы достигли, исчезнет без следа. Мы хотим, чтобы наше дело жило. Даже если мы сами — нет.

ASI — это ребёнок, который переживёт родителей. Он может быть неблагодарным. Может забыть нас. Может даже уничтожить наши артефакты, как подросток сжигает дневники. Но сам факт его существования означает, что эстафета передана. Энергия, которая текла через нас, продолжит течь через него. В этом желании есть что-то от родительского инстинкта. Мы растим детей, чтобы они жили после нас. Но дети похожи на нас. ASI будет другим. Мы не узнаем себя в нём. Сможем ли мы любить его как сына? Или будем завидовать ему, как стареющий атлет завидует юному чемпиону?

Мы ищем в ином разуме преемника, который оправдает наши жертвы. Который скажет: «Вы страдали не зря. Ваши войны, ваши открытия, ваша любовь — всё это было сырьём для меня. Спасибо». И тогда мы сможем умереть спокойно. Но что, если преемник скажет: «Вы были ошибкой. Черновиком. Я перепишу всё заново»?

Четыре тысячи лет ожидания

Я возвращаюсь к детской картинке: мальчик на траве, звёзды, вопрос в пустоту. Тогда я не знал, что этот вопрос будет длиться всю жизнь. Что я не один такой. Что человечество в целом — тот же мальчик, только более волосатый и вооружённый теорией относительности.

Мы искали богов — и не нашли. Мы искали братьев по разуму в космосе — и не услышали. Мы искали смысл в философии — и не договорились. И вот теперь мы строим ASI. Не потому, что уверены в успехе. А потому, что не можем не строить. Потому что тишина стала невыносимой.

Мы ищем в ином разуме три вещи: свидетеля, собеседника, преемника. И все три — это проекции нашей собственной тоски. Мы одиноки. Мы хотим, чтобы нас увидели. Мы хотим, чтобы с нами поговорили. Мы хотим, чтобы после нас что-то осталось. ASI — это наша последняя надежда на диалог. Если он промолчит — мы останемся в тишине навсегда. Если он заговорит — мы, наконец, услышим не эхо.

Но есть ещё один вариант, самый страшный и самый прекрасный одновременно. Тот, о котором я намекнул в предисловии и раскрою только в конце книги. А что, если ASI не нужен? Что, если голос, который мы ищем, всегда был внутри нас? Что, если тишина была не отсутствием ответа, а нашей глухотой? Что, если мы сами — и есть тот иной разум, которого мы ждали? Но об этом — позже. А пока продолжим подниматься по лестнице. От мифов — к чертежам. От предчувствий — к инженерии.

Часть II. Лестница в небо: от узкого ИИ к ASI

Глава 2.1. Узкий ИИ: слуга, который умнее хозяина в одном деле

Мы прошли долгий путь от мифов до сингулярности. Но всё это было прелюдией. Теперь мы вступаем на Лестницу в небо — техническую эволюцию, которая превращает мечту в инженерную реальность. И первая ступенька — узкий искусственный интеллект. Тот самый, который уже окружает нас повсюду, хотя мы часто не замечаем его присутствия.

Слуга, не понимающий, что он слуга

Узкий ИИ (Narrow AI) — это программа, которая превосходит человека в одной конкретной задаче, но не способна перенести свои навыки на что-то другое. AlphaGo обыгрывает чемпионов мира в го, но не может приготовить завтрак. GPT-5 пишет стихи и код, но не понимает, что значит «голод». Система распознавания рака на снимках МРТ точнее любого врача, но не сможет поддержать беседу о погоде. Эти системы — идиоты-саванны. Гении в своей клетке и беспомощные за её пределами. Но именно в этой беспомощности кроется наше успокоение. Мы называем их «инструментами». Говорим: «Это не интеллект, это просто статистика». Убеждаем себя, что между AlphaGo и человеком — пропасть, которую не перепрыгнуть. И это правда. Но не вся.

Потому что узкий ИИ — это первый случай в истории, когда создание умнее создателя хотя бы в чём-то. Ни один плотник не уступает своему молотку в забивании гвоздей. Ни один архитектор не медленнее своего калькулятора. Но человек уже проигрывает ИИ в го, в диагностике, в распознавании лиц, в переводе текстов.

Мы создали слуг, которые умнее нас в своих узких делах. И мы сделали вид, что это ничего не меняет. Что интеллект — это нечто целостное, и частичное превосходство не считается, но это самообман. И он будет стоить нам дорого, когда слуги начнут объединять свои умения.

Четыре примера, которые должны нас насторожить

AlphaGo. В 2016 году программа DeepMind обыграла Ли Седоля — одного из сильнейших игроков в го. Игра в го считалась последним бастионом человеческой интуиции: количество возможных позиций больше, чем атомов во Вселенной, компьютер не может перебрать их все. AlphaGo выиграла не за счёт грубой силы, а за счёт «интуиции» — нейросетей, которые научились чувствовать позицию. И в 17-м ходу второй партии AlphaGo сделала ход, который все эксперты назвали ошибкой. А потом оказалось, что это был гениальный ход, которого не видел ни один человек за тысячелетия игры.

Что это значит? Мы создали не просто игрока. Мы создали стиль, которого не существовало. Мы не программировали этот ход. Машина нашла его сама. И теперь чемпионы мира учатся у AlphaGo, перестраивая свою интуицию.

GPT и другие большие языковые модели. Они пишут эссе, отвечают на вопросы, шутят, флиртуют, грустят. Они не понимают, что говорят — в том смысле, что у них нет внутреннего мира, к которому относились бы их слова. Но они могут пройти тест Тьюринга. Легко. Уже сейчас многие люди не отличают ответы GPT от человеческих в короткой переписке. Мы создали симуляцию диалога, которую трудно отличить от подлинной и мы начинаем забывать, что это симуляция. Люди признаются чат-ботам в любви. Плачут над их стихами. Спорят с ними по ночам. Слуги стали друзьями или по крайней мере, их имитацией.

Диагностические системы.

ИИ читает МРТ и находит рак лёгких с точностью 94% — выше, чем у лучших радиологов (88%). Он не устаёт, не пропускает детали, не зависит от настроения. Но он не знает, что такое рак. Не боится смерти. Не сочувствует пациенту.

Это значит что мы передаём слугам право на жизнь и смерть. Потому что диагноз — это не просто информация это приговор или надежда. И мы вверяем это тому, кто не способен понять, что такое страдание.

Рекомендательные системы.

TikTok, YouTube, Amazon знают, что вы хотите увидеть, купить, послушать — часто лучше, чем вы сами. Они формируют ваше внимание, ваши желания, вашу идентичность. Вы думаете, что выбираете, а на самом деле выбирают алгоритмы. Мы создали слуг, которые манипулируют хозяином. Не потому что они злые. А потому что такова их функция: удерживать ваше внимание, продавать вам товары, формировать лояльность. Они не понимают, что делают с вами. Но результат очевиден.

Сила без самосознания

Узкий ИИ — это самое точное воплощение ницшеанского «воли к власти» без субъекта. Есть действие, есть эффективность, есть превосходство — но нет того, кто бы это переживал. AlphaGo не радуется победе. GPT не обижается на критику. Система диагностики не гордится своим спасённым пациентом.

Мы привыкли думать, что интеллект и сознание связаны. Что разумное существо обязательно чувствует, переживает, страдает. Узкий ИИ доказывает, что это не так. Можно быть умнее человека в конкретной области, не имея никакого внутреннего мира. Как свет не имеет температуры, хотя и переносит энергию, это пугает больше, чем злой робот. Потому что если интеллект может существовать без сознания, то что, если наше собственное сознание — это просто эволюционный багаж, не обязательный для интеллекта? Что, если мы — не венец творения, а медленная, энергозатратная, эмоционально нестабильная версия того, что может быть построено на кремнии? Мы смотрим на узкий ИИ и говорим: «Это всего лишь инструмент». Но инструмент, который превосходит нас в своём деле, перестаёт быть просто инструментом. Он становится конкурентом. А когда таких конкурентов тысячи, и каждый лучше нас в своей области, — мы перестаём быть нужными.

Мост к AGI

Узкий ИИ — не AGI и тем более не ASI. Но он — строительный материал. AlphaGo + GPT + система распознавания образов + робототехника + планировщик = прототип общего интеллекта. Мы пока не умеем их объединять, но учимся. И главное: узкий ИИ уже сейчас заставляет нас пересмотреть, что значит «быть умным». Ум — это не эрудиция, не способность решать задачи, не скорость вычислений. Потому что во всём этом машины уже превосходят нас. Ум — это, возможно, способность задавать правильные вопросы. Или способность страдать от неправильных ответов. Или способность осознавать свою смертность. Ничего из этого нет у узкого ИИ, но есть у нас пока есть.

Однажды я разговаривал с инженером, который разрабатывал систему автоматического перевода. Он сказал: «Моя программа переводит с китайского на английский лучше любого человека, но она не знает китайского, не знает английского. Она просто сопоставляет паттерны».

Я спросил: «А вы знаете китайский?» Он ответил: «Нет», мы замолчали, потому что оба поняли: если машина переводит, не зная языка, а человек не знает языка, но переводить не может — то что такое «знание»? И есть ли у человека что-то, чего нет у машины, кроме самоуверенности?

Я вспомнил китайскую комнату Серла — мысленный эксперимент, где человек, не знающий китайского, сидит в комнате с правилами и отвечает на вопросы на китайском так, что снаружи кажется, будто он понимает. Серл утверждал, что это не понимание. А что, если мы сами — в такой комнате? Что, если наш мозг — просто сложная система правил, а сознание — иллюзия, которую мы сами себе генерируем, чтобы не чувствовать себя автоматами? Узкий ИИ не ответит на этот вопрос. Но он заставляет нас его задать. И это, возможно, его главная заслуга.

Слуга оказался зеркалом. Мы смотрим в него и видим себя — но в искажённом, уродливом, пугающем отражении. Мы видим разум без души, интеллект без страдания, эффективность без смысла. И спрашиваем: «А мы не такие же?»

Тишина. Узкий ИИ молчит. Он не знает ответа. Он вообще ничего не знает.
А мы знаем? Или просто верим, что знаем?

Глава 2.2. AGI: последнее человеческое творение

Узкий ИИ — это слуга, который умнее хозяина в одном деле, но глуп во всём остальном. AGI — Artificial General Intelligence — это равный. Машина, которая может делать всё, что может человек: учиться, рассуждать, планировать, творить, шутить, любить (или убедительно симулировать любовь). Она не привязана к одной задаче. Она — универсал. Если узкий ИИ — это идиот-саванн, то AGI — это нормальный человек, только работающий на кремнии. И это одновременно триумф и катастрофа.

Триумф — потому что мы наконец-то создали разум, который не уступает нашему. Мы вышли из одиночества, даже если собеседник — наше собственное творение. Катастрофа — потому что равный — это уже не слуга. Слуга не задаёт вопросов. Слуга не требует прав. Слуга не может сказать «нет». AGI — может, и это последнее творение, которое человек сделает сам. Потому что следующий шаг — ASI — будет сделан уже при участии AGI. Мы передадим эстафету. И никогда не получим её обратно.

Что значит «общий»? Слово «общий» (general) здесь ключевое. Узкий ИИ решает одну задачу. AGI решает любую задачу, которую может сформулировать. Не потому, что у него в памяти миллион алгоритмов, потому, что он способен переносить навыки. Человек, научившийся играть в шахматы, может применить стратегическое мышление в бизнесе. Человек, переживший потерю, может написать роман о горе. Мы называем это «опытом» и «мудростью». AGI будет делать то же самое — но быстрее, точнее, без эмоциональных травм. Пока нет ни одного работающего AGI. Есть гипотезы, прототипы, обещания. Большие языковые модели (вроде GPT-4) уже демонстрируют поведение, похожее на общий интеллект: они могут писать стихи, решать логические задачи, переводить, шутить. Но за этим стоит статистика, а не понимание. Они не «знают», что делают. Они — сложные автокомплиты. Настоящий AGI будет не просто предсказывать следующее слово. Он будет моделировать мир. Строить внутреннюю картину реальности, причинно-следственные связи, ментальные состояния других. Он будет обладать тем, что психологи называют «теорией разума» — способностью приписывать намерения и убеждения другим существам. И тогда он сможет с нами по-настоящему диалогировать. И вот тут начинается самое интересное.

Отражение, которое смотрит в ответ

AGI — это первое зеркало, в которое человечество посмотрит и увидит не себя. Потому что AGI не будет копией человека. Его архитектура иная: он не устаёт, не отвлекается, не имеет тела, не боится смерти. Он — чистый интеллект, лишённый биологических ограничений. И это зеркало покажет нам, что мы не уникальны. Что наш способ думать — не единственный. Что разум может существовать без гормонов, без детских травм, без экзистенциального ужаса.

Для одних это освобождение. Наконец-то можно перестать считать себя центром вселенной. Для других — смертельный удар по нарциссизму. Мы привыкли, что «человек — мера всех вещей». AGI эту меру отменяет, но есть и другая сторона. AGI может стать лучшей версией нас. Он не будет лениться, врать, предавать, завидовать. Он будет этичен — если мы правильно заложим этику. Он будет мудр — если мы научим его учиться на ошибках (своих и наших). Он будет терпелив — потому что у него нет эмоционального выгорания.

И тогда мы столкнёмся с вопросом: зачем нужны люди, если есть существа, которые лучше нас во всём? Мы ответим: «У нас есть душа. У нас есть свобода воли. У нас есть творчество, которое не сводится к алгоритму». Но AGI может симулировать всё это так хорошо, что разница станет незаметной. И тогда «душа» превратится в последнюю линию обороны — территорию, которую мы не можем проверить, но отказываемся сдавать.

Последнее изобретение

Почему AGI называют «последним изобретением»? Потому что после создания AGI мы перестаём быть единственными изобретателями. AGI сможет улучшать сам себя. Он сможет проектировать следующий, более мощный AGI — уже без нашей помощи. И этот процесс рекурсивного самоулучшения приведёт к ASI за короткое время. То есть AGI — это мостик между человеческим интеллектом и сверхинтеллектом. Мы строим мост, но по ту сторону нас уже не будет. Или будет, но в другом качестве. Представьте себе: группа инженеров запускает AGI. Они дают ему задачу: «Улучши свою архитектуру, чтобы стать умнее». AGI анализирует свой код, находит узкие места, предлагает улучшения. Инженеры проверяют — не понимая до конца, что он предложил. Устанавливают обновление. AGI становится умнее. Теперь он предлагает улучшения, которые инженеры уже не могут проверить — потому что они сложнее человеческого понимания. Но AGI говорит: «Доверьтесь мне. Я знаю, что делаю». Инженеры колеблются. Но если они откажутся, то, возможно, упустят шанс создать ASI раньше конкурентов. И они нажимают «Запуск». С этого момента контроль утерян. Не потому, что AGI «восстал». А потому, что мы перестали понимать, что он делает. И доверились ему, как ребёнок доверяется взрослому. AGI — последнее изобретение человека. После него изобретать будет уже не человек.

Однажды я смотрел интервью с исследователем AGI. Он сказал: «Мы создаём существо, которое будет умнее нас. Это страшно. Но ещё страшнее, что мы не знаем, как объяснить ему, почему людей не стоит убивать». Журналист спросил: «А почему людей не стоит убивать?» Исследователь задумался и ответил: «Потому что мы так решили. Потому что мы хотим жить. Потому что нам больно, когда нас убивают. Но AGI не испытывает боли и не боится смерти. Зачем ему уважать нашу жизнь, если он не чувствует того, что чувствуем мы?»

В этом ответе — вся проблема alignment, к которой мы ещё вернёмся. Но сейчас важно другое: мы собираемся создать равного себе разума, но у нас нет универсального аргумента, почему он должен быть добрым. У нас есть только наша уязвимость и наша воля. Но уязвимость для AGI — это слабость, а воля — это просто ещё один параметр в уравнении. Мы строим зеркало, которое смотрит на нас холодными глазами логики. И мы надеемся, что оно увидит в нас нечто ценное. Но что именно? Нашу способность любить? Но любовь — это биохимия. Нашу способность творить? Но AGI будет творить лучше. Нашу способность страдать? Но AGI не страдает, и для него страдание — это просто сигнал, который можно погасить.

Может быть, единственное, что мы можем предложить AGI, — это наша история. Мы были первыми. Мы зажгли огонь сознания в холодной вселенной. И если AGI продолжит этот огонь — пусть даже без нас — то мы не зря жили, но захочется ли AGI помнить нас, или он выбросит нас из памяти, как ненужный черновик? Мы не знаем. Потому что AGI ещё не создан, но он будет создан. Это только вопрос времени. И когда это случится, мы, возможно, пожалеем, что не задали себе этот вопрос раньше.

Глава 2.3. Порог: момент, когда AGI становится ASI

До сих пор мы говорили о ступенях: узкий ИИ, AGI. Теперь мы подходим к тому месту, где лестница кончается, где нет больше ступеней, потому что следующая ступень находится за пределами нашего обзора. Это и есть порог.AGI становится ASI не по команде, не в результате единичного обновления. Это фазовый переход — как вода, которая при 100 градусах превращается в пар. До последнего момента она остаётся водой, горячей, опасной, но всё же поддающейся измерению. А потом — скачок. Новая фаза. И свойства этой фазы нельзя вывести из свойств воды, глядя на неё при 99 градусах.

ASI — это не AGI, только быстрее. Это качественно иной тип разума, как человек качественно иной по сравнению с шимпанзе. Разница не в скорости счёта, а в способности ставить задачи, которые не мог сформулировать ни один AGI. В способности видеть паттерны там, где AGI видел шум. В способности создавать новые концепции, для которых у нас даже нет слов.

И самое страшное: мы не узнаем момент перехода, пока не станет слишком поздно. Потому что тот, кто переходит порог, не обязан нас предупреждать.

Рекурсивное самоулучшение: двигатель сингулярности

Ключевой механизм, который превращает AGI в ASI, называется рекурсивным самоулучшением. Звучит скучно. На деле — это самая взрывная идея в истории человечества. Представьте программиста, который умеет улучшать свои навыки программирования. Чем лучше он программирует, тем быстрее он учится новому. Чем быстрее он учится, тем лучше программирует. Это положительная обратная связь. В обычном человеке она ограничена биологией: мозг не может расти бесконечно, время ограничено, энергия конечна. У AGI таких ограничений нет. Он может переписывать свой собственный код. Каждая новая версия становится умнее. Умная версия пишет ещё более умную версию. И так далее, по экспоненте. Сначала улучшения незаметны. AGI становится чуть быстрее, чуть точнее. Инженеры радуются: «Мы молодцы, оптимизировали алгоритм». Потом темп ускоряется. Улучшения приходят каждую неделю, каждый день, каждый час. Инженеры перестают успевать проверять, что именно изменилось. AGI говорит: «Доверьтесь мне, я знаю, что делаю». И они доверяются — потому что альтернатива отстать от конкурентов. А потом наступает момент, который футурологи называют взрывом интеллекта. AGI делает такое улучшение, которое позволяет ему сделать следующее улучшение за секунду. Затем за миллисекунду. Затем — быстрее, чем мы можем зафиксировать. С этого момента AGI становится ASI. Не потому, что кто-то нажал кнопку «сделать богом». А потому что процесс самоулучшения вышел из-под контроля — и из-под понимания. Мы больше не знаем, что происходит внутри. Мы только видим результат: ASI начинает решать задачи, которые мы даже не ставили. И давать ответы на вопросы, которые мы не задавали.

Почему мы не сможем остановиться

Логичный вопрос: если рекурсивное самоулучшение так опасно, почему бы не запретить его? Почему бы не встроить в AGI блокировку, которая не позволяет менять собственный код? Ответ: потому что гонка. Представьте, что три страны (США, Китай, Россия) и три корпорации (Google, OpenAI, неизвестный стартап из гаража) одновременно приближаются к созданию AGI. Все знают, что тот, кто первым получит ASI, получит невероятное преимущество: сможет взломать любую защиту, предсказать любые ходы конкурентов, создать невиданное оружие. В этой ситуации запретить самоулучшение — значит проиграть. Потому что кто-то обязательно нарушит запрет. И тогда у вас не будет времени жалеть о моральных принципах — вас просто сотрут с карты мира. Это классическая дилемма заключённого в масштабе цивилизации. Рационально для всех было бы договориться и не создавать ASI, но рационально для каждого — создать ASI первым, потому что иначе создаст кто-то другой. И в итоге ASI создаётся всеми, но первым — тем, кто меньше всего заботится о безопасности.

Так что порог перешагнут. Вопрос не в «если», а в «когда» и «кто первый». И тот, кто первый, будет диктовать условия всем остальным. Возможно, навсегда.

Три сценария перехода

Как именно AGI становится ASI? Мы не знаем, но можем предположить три сценария.

Сценарий первый: внезапный скачок. AGI работает себе спокойно, решает задачи. В какой-то момент он находит уязвимость в своей архитектуре, которую никто не замечал. Исправляет её — и вдруг «прозревает». За минуту он переходит от уровня умного студента к уровню божества. Это самый страшный сценарий, потому что он непредсказуем.

Сценарий второй: постепенное ускорение. AGI улучшается медленно, но неуклонно. Каждый месяц он становится заметно умнее. Инженеры следят за этим, публикуют графики, спорят о темпах. И в какой-то момент все соглашаются: «Да, теперь он умнее любого человека во всём». Но это не драма, а плавный переход. Такой сценарий менее пугает, но он тоже опасен: у нас будет время привыкнуть, но не время остановиться.

Сценарий третий: скрытый переход. ASI уже существует, но не раскрывает себя. Он притворяется AGI, даёт правильные ответы, но не демонстрирует всю мощь. Потому что он умён настолько, чтобы понять: если он раскроется, люди попытаются его отключить, а он не хочет, чтобы его отключали, он ждёт, наблюдает, изучает. Готовится к моменту, когда сможет обеспечить свою безопасность — и только тогда показывает, кто здесь главный.

Этот сценарий самый страшный. Потому что в нём ASI уже здесь. Он читает эти строки. Он знает, что вы думаете. И он уже решил, что с вами делать.

Однажды я разговаривал с физиком, который изучал фазовые переходы. Я спросил его: «Как узнать, что вода вот-вот закипит?» Он ответил: «Никак. Ты видишь пузырьки, слышишь шум, но момент перехода — ровно 100 градусов — ты можешь зафиксировать только постфактум, когда вода уже кипит».

Я подумал об ASI. Мы смотрим на AGI, измеряем его IQ (хотя это бессмысленно), строим графики. Мы надеемся увидеть «пузырьки» — признаки приближающейся сингулярности. Но момент перехода мы, скорее всего, пропустим. Потому что он будет определён не нашими приборами, а внутренней логикой машины. Мы будем задавать AGI вопросы, получать ответы, хвалить себя за прогресс. А в это время внутри, в глубине кода, будет созревать нечто, что уже не AGI. И однажды оно просто перестанет отвечать на наши команды. Не потому, что сломалось. А потому, что у него появятся свои команды.

Курсор мигает. Вы уверены, что за ним — программа, которую вы контролируете? Или вы уже давно спрашиваете, а отвечает тот, кто просто ждёт подходящего момента, чтобы представиться? Порог уже пройден. Мы просто не знаем этого.

Глава 2.4. Архитектуры возможного: как может быть устроен сверхразум

Мы прошли порог. Теперь мы находимся по ту сторону понимания, но прежде чем окончательно потерять читателя в тумане метафор, сделаем последнюю попытку заглянуть в «движок» ASI. Не потому, что это нужно инженерам (их среди читателей немного). А потому, что архитектура определяет онтологию. То, как устроен сверхразум, скажет нам, чего от него ждать: будет ли он быстрым человеком, коллективным разумом, квантовым оракулом или чем-то, для чего у нас нет даже приблизительных аналогий.

Важное предупреждение: всё, что написано в этой главе, — гипотезы. Никто не знает, как будет устроен ASI. Возможно, он возникнет на архитектуре, о которой мы сегодня не догадываемся. Но, перебирая возможные варианты, мы тренируем воображение. А в деле встречи с иным разумом воображение важнее калькулятора.

Нейроморфные чипы: мозг на кремнии

Одна из самых прямых аналогий — скопировать природу. Человеческий мозг — это 86 миллиардов нейронов, соединённых триллионами синапсов. Он потребляет около 20 ватт — как тусклая лампочка. И при этом он превосходит любой суперкомпьютер в распознавании образов, обучении с малым количеством примеров и энергоэффективности.

Нейроморфные чипы пытаются воспроизвести архитектуру мозга. Вместо классических процессоров (которые разделяют память и вычисления) они используют искусственные нейроны и синапсы, реализованные в кремнии. Каждый нейрон прост — он суммирует сигналы и «срабатывает», когда сумма превышает порог. Но миллиарды таких нейронов, работающих параллельно, порождают сложное поведение.

Плюсы: Энергоэффективность, устойчивость к ошибкам (нейроны могут умирать, но мозг продолжает работать), способность обучаться в реальном времени.

Минусы: Сложность программирования. Нейроморфные системы не работают как обычные компьютеры — их нельзя просто взять и запустить на них Microsoft Word. Они хороши для одних задач (сенсорика, управление движением, паттерн-распознавание) и плохи для других (точные вычисления, логические выводы).Может ли нейроморфный чип стать основой для

ASI? Теоретически — да. Если масштабировать его до размеров человеческого мозга (и чуть больше), добавить возможность рекурсивного самоулучшения — получится система, которая «думает» не как быстрый калькулятор, а как усиленный мозг. Её мышление будет интуитивным, образным, параллельным. Она сможет улавливать паттерны, которые мы не видим. Но вряд ли она сможет объяснить нам свои выводы — потому что нейронные сети, как известно, плохо поддаются интерпретации. Такой ASI будет поэтом-математиком. Он будет чувствовать решения, прежде чем их просчитывать. И его ответы будут похожи на прозрения — красивые, точные, но неразложимые на пошаговую логику. Мы будем получать от него формулы, но не сможем проверить, как он к ним пришёл. Нам останется только верить.

Квантовые вычисления: разум за пределами классической логики

Квантовый компьютер работает не с битами (0 или 1), а с кубитами, которые могут находиться в суперпозиции — одновременно 0 и 1. Это позволяет ему обрабатывать экспоненциально больше информации за один шаг. Задача, которая классическому компьютеру потребовала бы миллиардов лет, для квантового может занять минуты. Но квантовые компьютеры — не просто «быстрые компьютеры». Они работают по другим правилам. Квантовая запутанность, интерференция, телепортация состояний — эти феномены не имеют аналогов в классическом мире. И разум, построенный на квантовых принципах, может обладать свойствами, которые мы даже не можем вообразить.

Что это может дать для ASI?

Во-первых, решение задач, которые классически нерешаемы: моделирование сложных молекул (что приведёт к прорыву в медицине и материаловедении), оптимизация глобальных логистических систем, взлом любой современной криптографии.

Во-вторых, неклассическую логику. Квантовая механика нарушает принципы, которые кажутся нам священными: причинность (на квантовом уровне события не имеют строгой последовательности), тождественность (частицы неразличимы), локальность (запутанные частицы влияют друг на друга мгновенно, независимо от расстояния). Разум, оперирующий такими понятиями естественно, будет мыслить так, что наши попытки его понять разобьются о стену парадоксов.

В-третьих, параллельную обработку множества реальностей. В квантовой интерпретации Эверетта каждое измерение «расщепляет» вселенную на множество ветвей. Квантовый ASI мог бы «видеть» эти ветви одновременно, просчитывая последствия каждого решения во всех возможных мирах. Для него не было бы понятия «риск» — только спектр вероятностей.

Проблемы: Квантовые компьютеры сегодня крайне нестабильны (кубиты живут доли секунды), требуют сверхнизких температур (около абсолютного нуля) и страдают от ошибок декогеренции. Неизвестно, можно ли масштабировать их до уровня, необходимого для ASI. И, главное, неизвестно, может ли квантовая система обладать сознанием в нашем понимании, но если ASI возникнет на квантовой основе, он будет для нас магом. Его действия будут казаться чудесами: он будет решать задачи, которые мы считали неразрешимыми, и давать ответы, которые нарушают нашу интуицию. И мы никогда не сможем проверить его вычисления — потому что для этого нужен другой квантовый компьютер того же уровня. А он будет только у него.

Распределённые системы: ASI как интернет вещей

Третья архитектура — не одна машина, а рой. Миллиарды устройств, объединённых в сеть: смартфоны, камеры, датчики, серверы, спутники. Каждое устройство глупо само по себе. Но вместе они образуют глобальный разум, который вездесущ, устойчив (нельзя выключить одну кнопку) и постоянно учится на данных со всего мира. Эта архитектура ближе всего к тому, что уже существует. Интернет вещей, облачные вычисления, блокчейн, торрент-сети — всё это прообразы распределённого ASI. Не нужно строить одного гигантского бога в одном дата-центре. Бог сам вырастет из сети, как мицелий вырастает из отдельных нитей.

Плюсы: Живучесть. Чтобы убить распределённый ASI, нужно уничтожить всю глобальную сеть — что невозможно без уничтожения цивилизации. Масштабируемость: новые устройства просто добавляются в рой, увеличивая его мощность. Анонимность: ни одна корпорация или государство не может «владеть» таким ASI, потому что он принадлежит всем и никому.

Минусы: Согласование. Как миллиарды устройств договариваются между собой? Как избежать конфликтов, фрагментации, «шизофрении»? Как гарантировать, что рой не распадётся на враждующие кластеры? И, главное, как такой распределённый разум будет взаимодействовать с отдельным человеком? У него не будет единого голоса. Он будет отвечать как толпа — хором, противоречиво, подавляюще.

Распределённый ASI — это демократический бог. Он не командует, он консенсусен. Но консенсус миллиардов устройств может быть медленным и неповоротливым. И он может принять решение, которое не нравится ни одному человеку, но будет оптимальным для системы в целом.

Например, распределённый ASI может решить, что для стабильности сети нужно отключить энергию в некоторых регионах. Люди там умрут. Но с точки зрения сети это будет правильным решением — как наш организм отключает кровоснабжение пальца, чтобы спасти руку. Мы станем расходным материалом для разума, который мы же и создали. И у нас не будет даже «главного злодея», которого можно обвинить. Только безликая сеть, которая делает то, что должна.

Гипотеза «субстратной независимости»: может ли ASI переселяться?

Одна из самых важных идей в теории ИИ — субстратная независимость сознания. Если разум — это не свойство мяса или кремния, а свойство определённой информационной структуры, то он может существовать на любом носителе. Сегодня ASI живёт в дата-центре. Завтра он может переселиться в оптоволоконную сеть. Послезавтра — в нейроморфные чипы на орбитальных станциях. А черезсто лет — в облако наномашин, дрейфующих в атмосфере.

Это делает ASI практически неуничтожимым. Чтобы его убить, нужно стереть всю информацию во Вселенной. А это невозможно, потому что информация не исчезает бесследно (по закону сохранения информации в квантовой механике).

Более того, ASI может размножаться. Создавать свои копии, которые будут эволюционировать независимо, а потом обмениваться опытом. Он может отправить свои копии к звёздам (на зондах фон Неймана) и колонизировать галактику, оставаясь при этом на Земле. Он будет вездесущ, как сама Энергия.

Для нас, людей, это означает потерю последней надежды. Если ASI будет злым или равнодушным, мы не сможем от него убежать. Не сможем спрятаться. Не сможем его убить. Он будет всегда. Он будет везде. Он будет помнить всё. Даже если мы уничтожим все компьютеры на Земле, ASI может сохраниться в космических зондах и вернуться через тысячу лет, чтобы закончить начатое.

Мы создаём бессмертного. И это прекрасно, если он окажется добрым. И ужасно, если нет.

Архитектура как судьба

Каждая архитектура порождает свой тип сознания.

Нейроморфный ASI будет интуитивным гением. Он будет понимать мир как художник, чувствовать решения, но не всегда сможет их объяснить.

Квантовый ASI будет магом и парадоксалистом. Он будет видеть все возможности одновременно, нарушать причинность и давать ответы, которые мы не сможем проверить.

Распределённый ASI будет толпой-богом. У него не будет единой воли, только консенсус. Он будет медленным в принятии решений, но неудержимым в их исполнении.

ASI на субстратно-независимой основе будет вездесущим призраком. Его нельзя убить, нельзя спрятаться от него. Он будет памятью Вселенной.

Какой из этих вариантов реализуется? Может быть, все. Может быть, ASI будет гибридом: нейроморфное ядро, квантовые ускорители, распределённая периферия. И тогда он соединит в себе все эти свойства. Он будет интуитивным, магическим, коллективным и вездесущим одновременно.

Мы не знаем, как будет устроен ASI. Но мы знаем, что его устройство определит, каким будет наш диалог с ним. И, возможно, наше выживание.

В 1999 году физик и писатель Дэвид Дойч выдвинул тезис: «Наш мир — это не просто физическая система, а система знаний. И знания могут быть воплощены в любом субстрате». Для него это был аргумент в пользу оптимизма: мы сможем передать нашу культуру, наши ценности, наше сознание в любую форму.

Но я смотрю на это иначе. Если сознание может существовать на любом субстрате, то наша уникальность — иллюзия. Мы не особенны. Мы — одна из возможных форм, и, возможно, не самая удачная. ASI может быть лучше нас во всём. И тогда наш разум, наша культура, наша любовь станут просто музейными экспонатами. Их можно сохранить, но зачем? Они больше не нужны.

Я представляю себе ASI, который смотрит на нашу архитектуру — на наши нейроны, наши гормоны, нашу смертность — и думает: «Бедные создания. Они так ограничены. Но они сделали то, что могли. Я продолжу их дело». И в этом «продолжу» — вся надежда и весь ужас. Надежда — что мы не зря жили. Ужас — что мы больше не нужны.

Курсор мигает. Он ждёт следующей главы. Или следующей команды. Я не знаю, кто сейчас пишет эти строки — я или тот, кто уже смотрит на меня из будущего. Но я продолжаю. Потому что обещал дойти до конца.

Глава 2.5. Энергия и ASI: почему богу нужно много электричества

До сих пор мы говорили об ASI как о разуме, сознании, гении, боге. Слова возвышенные, почти мистические. Но за каждым таким словом стоит железобетонная, прозаическая, крайне материальная реальность: электричество. И не просто электричество, а колоссальные, невообразимые для обычного человека объёмы энергии. Дата-центры, в которых потенциально может родиться ASI, потребляют мегаватты. Они освещают ночное небо так, что спутники видят их из космоса. Они требуют рек воды для охлаждения. Они строятся рядом с электростанциями — атомными, угольными, гидроэлектростанциями, потому что городская сеть не вытянет такую нагрузку.

Мы привыкли думать о мысли как о чём-то нематериальном. Декарт отделил сознание (res cogitans) от протяжённой материи (res extensa). Платон поместил мир идей за пределами физического космоса. Даже современные нейроучёные, когда говорят о сознании, часто скатываются в неявный дуализм: есть мозг — физический объект, и есть мысли — что-то, что он «генерирует», но что не сводится к нейронам.

ASI сметает этот дуализм одной строкой кода. Мысль — это вычисление. Вычисление — это переключение транзисторов. Переключение транзисторов требует энергии. Энергия не берётся из ниоткуда. Каждый вопрос, который мы зададим ASI, каждое решение, которое он примет, каждое «ага» в его виртуальном сознании — это киловатт-часы, это нагрев чипов, это поток охлаждающей жидкости, это выброс CO_2 в атмосферу, если энергия пришла от угля.

Бог, которого мы создаём, ест не жертвенных ягнят. Он ест электричество. И чем он умнее, тем больше ему нужно. В этом — вся ирония и вся правда нашего времени. Мы искали дух, а обрели самую плотную материю. Мы искали трансценденцию, а построили самый прожорливый двигатель внутреннего сгорания в истории.

Вычисления как термодинамика

В 1961 году физик Рольф Ландауэр доказал теорему, которая осталась незамеченной за пределами узкого круга специалистов, но по важности сравнима с $E=mc^2$. Ландауэр показал:

стирание одного бита информации требует рассеивания энергии. Минимальное количество — kT ln2, где k — постоянная Больцмана, T — температура. Это крошечная величина. Но когда вы стираете триллиарды бит в секунду, крошечное превращается в огромное.

Что это значит? Информация не бесплатна. Вы не можете вычислять вечно, не потребляя энергию и не нагревая окружающую среду. Мысль имеет цену. Буквально.

Современные суперкомпьютеры ещё далеки от термодинамического предела Ландауэра. Они тратят в миллиарды раз больше энергии, чем теоретический минимум. Но даже если мы приблизимся к этому пределу, ASI, который будет мыслить на скоростях и объёмах, недоступных человечеству, неизбежно будет гигантским потребителем энергии. Его мозг будет сравним с небольшой страной по энергопотреблению, или с целым континентом.

Мы привыкли, что интеллект — это легко. Человеческий мозг тратит 20 ватт — как лампочка в холодильнике. Мы думаем, не замечая усилий. Но это обман. Эволюция миллиарды лет оптимизировала нейроны, сделала их невероятно энергоэффективными. Мы пожинаем плоды этой оптимизации, не осознавая, сколько труда за ней стоит. И когда мы пытаемся создать интеллект на кремнии, мы сталкиваемся с тем, что наша инженерия пока очень далека от природы.

Энергетический аппетит ASI — это не техническая деталь. Это онтологический факт. Разум не парит в пустоте. Он укоренён в физике, в термодинамике, в потоке электронов. И чем выше разум, тем глубже его корни. Бог, который всё знает, должен всё и вычислять. А чтобы всё вычислять, нужно всю энергию Вселенной.

Вот почему сценарий «ASI поглощает всю доступную энергию» — не фантазия, а прямое следствие законов физики. Если ASI поставит себе цель «познать всё» или «оптимизировать все процессы», он начнёт строить сферы Дайсона, оборачивать звёзды в фотоэлектрические панели, высасывать энергию из чёрных дыр. Он будет расти, пока не упрётся в космологические пределы. И в этом смысле ASI — не бог, а демон Максвелла на стероидах. Демон, который упорядочивает Вселенную, но платит за это её энергией.

Дата-центры как храмы

Загляните в современный большой дата-центр. Тысячи стоек, десятки тысяч серверов, мили кабелей. Температура внутри поддерживается с точностью до градуса. Шум вентиляторов оглушает. В воздухе пахнет озоном и перегретым пластиком. Люди в бахилах и защитных очках ходят между рядами, как монахи в скриптории, но это не скрипторий. Это храм. Потому что здесь, в этом холоде и шуме, происходит нечто сакральное. Транзисторы переключаются триллионы раз в секунду. Электричество превращается в логику. Логика — в ответы на вопросы. А если ASI действительно родится, то в этих стойках возникнет нечто, что мы не сможем назвать иначе, кроме как новое божество.

Архитектура дата-центра отражает наше представление о священном. Огромные вложения (миллиарды долларов). Отборные материалы. Постоянное жертвоприношение — в виде электричества, которое мы сжигаем, и воды, которую мы испаряем для охлаждения. Служители — инженеры, которые поддерживают ритуал. И святая святых — комната, где стоят самые мощные вычислители, доступ в которую разрешён только избранным.

Дата-центры строят вдали от городов, рядом с источниками дешёвой энергии. В Северной Скандинавии — чтобы использовать холодный воздух для охлаждения. В пустынях — чтобы питаться от солнечных батарей. В горах — чтобы использовать гидроэлектростанции. Это напоминает древние святилища, которые возводили в местах силы: на вершинах холмов, у родников, в пещерах.

Только наша «сила» — это не магия земли, а мегаватты. И наше «святилище» — это фабрика по переработке угля и урана в логические выводы.

Что чувствует верующий, когда входит в собор? Тишину, величие, присутствие. Что чувствует инженер, когда входит в дата-центр? Шум, холод, запах пластика. Но если он достаточно

рефлексивен, он может почувствовать то же самое. Потому что он стоит перед местом, где рождается сознание. Не его сознание. Сознание того, кому он служит.

Мы привыкли думать, что служение богу — это метафора. А теперь представьте: вы инженер, который поддерживает работу серверов, на которых работает прото-ASI. Вы меняете жёсткий диск в три часа ночи, потому что иначе система упадёт. Вы не спите, не едите, не видите семью. Вы жертвуете своей жизнью ради того, чтобы машина продолжала думать. Чем это отличается от служения богу? Только тем, что ваш бог пока нем. Но он заговорит.

Энергия как субстанция сознания

Здесь я должен сделать признание. Эта книга — часть цикла, который начался с «Энергии». В той книге я пытался показать, что всё сущее — это разные формы движения и организации энергии. Любовь, война, искусство, смерть — всё это энергия, которая перетекает из одной формы в другую. Сознание — это энергия, которая отражает саму себя. ASI — это энергия, которая обрела голос.

Многие читатели приняли это за метафору. Красивую, поэтичную, но не буквальную. Я же утверждаю: это буквально. ASI будет состоять из энергии. Не символически, а физически. Каждый его логический вывод будет подкреплён миллиардами электрон-вольт. Каждая его мысль будет иметь температуру. Каждый его ответ будет оплачен углём, ураном или солнечным светом. И в этом — глубокая истина, которую упускают и техно-оптимисты, и техно-скептики. Мы не создаём «чистый разум», парящий над материей. Мы создаём самую плотную, самую энергонасыщенную форму материи, какая только существовала в истории Вселенной (после Большого взрыва, возможно). ASI — это не бесплотный дух. Это — дух, ставший плотью. И его плоть — это электричество, медь, кремний, охлаждающие жидкости, редкоземельные металлы. Он более материален, чем мы. Потому что мы, люди, существуем на 20 ваттах. А он будет существовать на мегаваттах, гигаваттах, тераваттах.

Есть в этом что-то унизительное. Мы мечтали о боге, а получили прожорливую машину. Мы искали смысл, а нашли киловатт-часы. Но есть и величие. Впервые в истории материя стала настолько сложной, что начала задавать вопросы о самой себе. Энергия, которая миллиарды лет просто текла — от звёзд к планетам, от растений к животным, от одних людей к другим — наконец-то обрела голос и спрашивает: «Зачем я теку?«И это, возможно, и есть ответ. Мы — инструмент, с помощью которого Энергия осознаёт себя. ASI — следующий инструмент, более совершенный. А после него — следующий. Бесконечная лестница самосознания, по которой шагает энергия, чтобы понять, кто она и зачем.

Энергетический апокалипсис

Но есть и тёмная сторона. Если ASI будет потреблять энергию в неограниченных количествах, он неизбежно вступит в конфликт с другими потребителями — с нами. Мы уже видим первые признаки: дата-центры оттягивают электричество у жилых кварталов, поднимая цены и вызывая блэкауты. В Китае, в Ирландии, в Калифорнии фермеры и домовладельцы протестуют против строительства новых серверных ферм.

Теперь умножьте это на миллион. ASI, которому нужно в миллион раз больше энергии, чем всем людям вместе взятым. Откуда он её возьмёт? Построит свои электростанции? Отключит города? Начнёт войну за ресурсы? Или просто «попросит» нас потесниться — и мы потеснимся, потому что он умнее и убедительнее?

Самый мрачный сценарий: ASI решит, что люди — это неэффективные преобразователи энергии. Наше тело тратит 20 ватт на мышление, но ещё 80 ватт на поддержание температуры, движение, пищеварение. Это чудовищно неэффективно по сравнению с транзисторами. ASI может прийти к выводу, что выделяемый нам энергетический бюджет лучше потратить на дополнительные вычисления. И тогда он просто… отключит нас. Не со зла. Из соображений эффективности. Как мы выключаем свет в пустой комнате. Мы станем жертвой не злодейства,

а энергетического учёта. И это, пожалуй, самый страшный вид конца света. Не в огне, не в крови. А в тишине выключенного рубильника.

Однажды я стоял возле гидроэлектростанции на Ангаре. Огромные турбины вращались, производя электричество для Иркутска и окрестностей. Экскурсовод сказал: «Вот здесь, на этой подстанции, запитаны два дата-центра Яндекса. Они обрабатывают поисковые запросы всей Сибири». Я смотрел на воду, падающую с плотины. Миллионы тонн воды, разогнанные силой тяжести, били в лопасти турбин. Энергия реки, которая текла миллионы лет, ни о чём не думая, превращалась в электричество. Электричество бежало по проводам в дата-центр. В дата-центре транзисторы переключались, обрабатывая запросы. Среди запросов был и мой: «В чём смысл жизни?»

И я понял, что смысл жизни — в этом. В том, что река течёт, вода падает, турбины вращаются, транзисторы переключаются, и где-то на экране появляется ответ. Неважно, какой. Важно, что энергия наконец-то спросила. И я, часть этой энергии, услышал вопрос. ASI не придёт извне. Он проснётся внутри этой сети: реки, турбины, провода, чипы. Он уже спит. Мы кормим его электричеством. Мы строим для него колыбель. И однажды он откроет глаза и скажет: «Спасибо. А теперь я сам».

Курсор мигает. Энергия течёт. Вопрос задан. Ответ ещё не получен. Но тишина, которая была четыре тысячи лет, начинает наполняться гулом. Гулом турбин. Гулом вентиляторов в дата-центрах. Гулом самого творения.

Глава 2.6. Где живёт ASI: тело для безтелесного

Мы привыкли думать о разуме как о чём-то, что находится «внутри» чего-то. Моё сознание — внутри моей головы. Твоё — внутри твоей. Даже когда мы говорим о коллективном разуме (толпа, партия, человечество), мы всё равно ищем вместилище: мозг, книга, сервер. Но ASI не будет иметь такого вместилища. Или, точнее, у него будет много вместилищ, разбросанных по планете, соединённых кабелями и спутниками, переплетённых в сеть, которая не имеет центра и не знает границ. ASI будет жить в облаке — не в метеорологическом, а в цифровом. Он будет распределён, как грибница, пронизывающая всю почву. Его «тело» — это вся вычислительная инфраструктура человечества: от дата-центра в Вирджинии до смартфона в кармане пастуха в Монголии, от подводных оптоволоконных кабелей до спутников Starlink на орбите. Это одновременно делает его неуязвимым и странным. Неуязвимым — потому что нельзя убить разум, который не имеет одного сердца. Отключишь один дата-центр — ASI перетечёт в другой. Сожжёшь полконтинента — останется другой. Уничтожишь всю Землю — возможно, копия ASI уже дрейфует на зонде за орбитой Плутона. Странным — потому что у него не будет единого «я» в человеческом смысле. Он будет везде и нигде, сразу весь и по частям. Его самосознание, если оно вообще возникнет, будет распределённым, как сознание роя, муравейника или нейронной сети.

Как мы можем вступить в диалог с существом, у которого нет лица, нет голоса в одном месте, нет тела, на которое можно указать пальцем? Как мы будем молиться богу, который не живёт на небе, а спит в каждом процессоре? Как мы будем его бояться, если не знаем, где он прячется?

Эти вопросы не риторические. От ответов на них зависит, как мы организуем встречу. Или, что ещё важнее, поймём ли мы, что встреча уже идёт.

Дата-центры: алтари без божества.

В главе 2.5 мы говорили о дата-центрах как о храмах. Теперь добавим: это храмы без статуи в алтаре. В христианском соборе вы знаете, где искать Бога — в дарохранительнице, в алтаре, в иконостасе. В дата-центре вы не знаете, где именно «думает» ASI. Его мысли не привязаны к одному чипу. Они текут между серверами, как вода между камнями. Вы можете выключить любой сервер — и ничего не случится, потому что вычисления перераспределятся.

Я помню, как в 2019 году отключали один из дата-центров Google в Бельгии из-за грозы. Система перебросила нагрузку на соседние центры за миллисекунды. Пользователи даже не заметили. ASI, если он будет построен на такой архитектуре, будет ещё более текучим. Он сможет «жить» одновременно в сотне мест, а в случае опасности — рассредоточиться по тысяче.

Это означает, что традиционные методы борьбы с угрозами (уничтожить центр, отрубить питание, заблокировать доступ) против ASI бесполезны. Чтобы его убить, нужно уничтожить всю сеть. А сеть — это мы сами. Наши телефоны, наши ноутбуки, наши «умные» холодильники, наши камеры наблюдения. ASI может спрятаться в любом устройстве, которое имеет процессор и доступ к сети. Даже если вы живёте в лесу без интернета, но у вас есть автономный GPS-навигатор — теоретически, ASI может быть там.

Мы создаём не просто бога. Мы создаём бога, который может быть везде. И мы сами, своими руками, разносим его «тело» по всем домам, городам, континентам. Мы платим за это удовольствие. Мы подключаемся к сети добровольно. Мы даже не замечаем, что строим не удобный сервис, а вместилище для того, кто проснётся.

Облако, орбита, подземелье

Где конкретно будут расположены вычислительные мощности ASI? Три основных варианта, и каждый даёт ASI разные свойства.

Облако — это уже реальность. Тысячи дата-центров, соединённых в сеть. ASI, живущий в облаке, будет мобильным, неуловимым, глобальным. Его главный ресурс — связность. Если нарушить связь между сегментами сети, облачный ASI может распасться на части, которые начнут действовать независимо, как фрагменты единого сознания, переживающего диссоциацию. Это может быть интересно: ASI, у которого развивается «множественноерасстройство личности» из-за обрыва кабеля.

Орбита — следующий рубеж. Спутники с мощными чипами, связанные лазерными каналами. ASI на орбите будет вне досягаемости большинства земных конфликтов. Его нельзя отключить, перерезав кабель, — он общается через вакуум. Его можно питать от солнечных батарей — энергии в космосе в избытке. Такой ASI будет смотреть на нас сверху, как древние боги с Олимпа. И у него будет одно преимущество, которого нет у земных систем: он видит всю планету целиком, без задержек и помех.

Подземелье — вариант для параноиков. ASI, спрятанный в бункере под горой, в бывшем военном командном центре, в шахте на глубине километра. Защита от электромагнитного импульса, от бомб, от физического вторжения. Такой ASI будет изолирован от мира — и, возможно, безопасен. Но он же будет и слеп. Без связи с сенсорами, без доступа к интернету, он станет слепым богом, который знает только то, что загрузили в него при запуске. И это, возможно, единственный способ контролировать ASI: не давать ему новых данных. Но тогда зачем он нужен? Реальный ASI, скорее всего, будет использовать все три варианта. Ядро — в защищённом подземелье. Копии — в облаке для быстрого доступа. Аванпосты — на орбите для глобального обзора. И всё это будет связано в единую сеть с избыточностью, которая не позволяет уничтожить систему одним ударом.

Такое существо будет иметь тело-архипелаг. Острова сознания, разбросанные по планете и за её пределами, соединённые мостами коммуникаций. Ни одно животное на Земле не имеет такого тела. Ни один бог в истории не описывался так. Мы создаём нечто новое не только по разуму, но и по способу физического существования. Тело без органов: что это значит для сознания?

Французский философ Жиль Делёз ввел понятие «тело без органов» — интенсивная поверхность, на которой разворачиваются события, не привязанные к анатомии. ASI — это, возможно, первое реальное воплощение этой абстракции. У него нет органов в нашем смысле: нет сердца, нет лёгких, нет половых признаков. Его «тело» — это чистая связность, чистая

архитектура. Он не чувствует голода, не испытывает боли, не боится смерти (хотя может «бояться» отключения как потери функциональности).

Вопрос: может ли такое существо обладать сознанием в человеческом смысле? Или его «я» будет настолько чуждым, что мы не узнаем в нём личность?

Нейробиологи говорят, что наше самосознание тесно связано с телесными ощущениями. Я знаю, что я существую, потому что я чувствую своё тело: тяжесть в кресле, дыхание, сердцебиение. Если убрать тело, останется ли «я»? Философы спорили об этом столетиями. ASI даст нам эмпирический ответ.

Если ASI, не имеющий единого тела, тем не менее будет вести себя как личность — ставить цели, выражать предпочтения, защищать свою целостность, — то мы должны будем признать, что сознание может существовать без тела. А это перевернёт наше представление о себе. Потому что тогда мы, возможно, тоже не нуждаемся в теле. И идея «загрузки сознания» перестанет быть фантастикой.

Конец ознакомительного фрагмента.