КОНЕЦ индивидуума

Путешествие философа в страну искусственного интеллекта

Гаспар Кёниг

Конец индивидуума. Путешествие философа в страну искусственного интеллекта

http://www.litres.ru/pages/biblio_book/?art=69553129
Конец индивидуума. Путешествие философа в страну искусственного интеллекта:
ISBN 978-5-6048294-3-1

Аннотация

Что ждет человека и его право на свободный выбор в век искусственного интеллекта? Чтобы ответить на этот насущный вопрос, французский философ Гаспар Кёниг предпринял кругосветное путешествие и познакомился с сотней ведущих специалистов по ИИ, хранящих ключи к будущему. Кёниг уверен, что страх перед потерей работы из-за распространения GPT-моделей – лишь верхушка айсберга: куда важнее, что они бросают вызов нашим представлениям о знании, творчестве и свободе воли. Безудержный рост ИИ уже несет нам власть без демократии, искусство без художника, экономику без рынка и справедливость без правосудия. Эта наполненная юмором и

надеждой книга поможет вам разобраться, какие возможности открывает человечеству эпоха ИИ и какие опасности она таит.

В формате PDF A4 сохранен издательский макет книги.

Содержание

Гаспар Кёниг
Конец индивидуума. Путешествие философа в страну искусственного интеллекта

Gaspard Kœnig

La Fin de l’individu Voyage d’un philosophe au pays de l’intelligence artificielle

L’OBSERVATOIRE

Предисловие

То, что моя книга выходит на русском языке в столь трагичный период европейской истории, кажется мне особенно важным. Я считаю, что вопреки глупости правящих элит мы должны поддерживать прямую связь между автором и читателем, связь человека с человеком, чтобы не утратить тонкую нить, соединяющую наши культуры и наши мысли.

Полевой этап моего исследования искусственного интеллекта стартовал в 2018 году. В такой динамичной области, где одна инновация стремительно сменяет другую, книга устаревает на следующий день после публикации. Преимущество философского подхода – в его аналитичности, и мои размышления пятилетней давности по-прежнему кажутся мне актуальными.

В качестве иллюстрации я предлагаю обратиться к ChatGPT, хотя это и означает несколько забежать вперед.

Появлению этого чрезвычайно успешного чат-бота (вероятно, первого прошедшего тест Тьюринга) сопутствуют две хорошо знакомые нам фобии, излюбленные прессой, падкой на сенсации.

Во-первых, страх, что так называемый сильный искусственный интеллект выработает самостоятельное сознание. Здесь следует вспомнить мысленный эксперимент философа Джона Серла с «китайской комнатой»: робота можно на-

учить делать вид, что он говорит по-китайски, обучив его реагировать на стимулы, связанные с теми или иными иероглифами. Но это лишь симуляция – он не будет говорить по-китайски. Другими словами, он не сможет понять смысл китайских слов. То же относится и к текстам ChatGPT – это лишь простая имитация человеческой мысли.

Вторая фобия – страх «конца работы». Однако ChatGPT, как и все ИИ, для выработки своих ответов опирается на результаты колоссального интеллектуального труда многих людей. Без этого сырья, без данных, собранных или проверенных реальными людьми (часто теми, чей труд плохо оплачивается и бывает временным), нет топлива для машины! Проблема заключается не в «конце работы», а в ее справедливой оплате, поэтому нам определенно придется заставить создателей ChatGPT выплачивать нам авторские гонорары.

Какими же вопросами тогда стоит задаваться?

В первую очередь теми, что угрожают нашим либеральным представлениям о науке и о свободе.

ChatGPT усиливает искушение доверить нашу свободу выбора машине. Мы уже можем положиться на алгоритмы, чтобы найти нужное жилье или даже партнера для отношений. Неужели теперь мы отдадим ей и самое сокровенное – наши мысли? Вопрос не в том, «разумен» ли ChatGPT, а в том, кто вообще имеет право на разум. На карту поставлено наше человеческое достоинство.

С технологической стороны ChatGPT бросает вызов меха-

низму производства знаний, который диктует современная наука. Основанные на бесчисленных корреляциях, рассуждения ChatGPT невоспроизводимы, неопровержимы, необъяснимы и, что особенно важно, не имеют источников. Сегодня впервые за 2500 лет прозвучала идея о том, что можно освободиться от концепта значимости источника. Уже в диалогах Платона появляется первый источник – Сократ. У античных историков источники, может быть, и не всегда надежны, но они есть: Тит Ливий тщательно цитирует других античных авторов, например Полибия, и приводит устные свидетельства. Сегодня академическая наука полагается, пусть иногда и чрезмерно, на качество справочного аппарата. В школе учителя стараются научить учеников правильно пользоваться интернетом – указывать сайты, на которые они ссылаются.

Конечно, если вы попросите ChatGPT предоставить свои источники, он сделает это. Но это не истинные источники, затерявшиеся в магме глубокого обучения, это всего лишь вероятные источники. Разница фундаментальна: неопределенность больше не считается дефектом, который нужно устранить, а воспринимается как структурная характеристика.

Такое пренебрежение источниками не относится к техническим ошибкам. Ответы ChatGPT основаны на корреляциях, которые по своей природе не поддаются аналитическому разложению. Они вдохновлены всеми правдами, неправ-

дами и квазиправдами, выплеснутыми в сеть. В этом смысле ChatGPT – полная противоположность «Википедии», которая при всех своих недостатках известна маниакальным пристрастием к источникам и пытается вывести объективную истину из дебатов между людьми-составителями.

Когда я писал эту книгу, мне встретилась статья Генри Киссинджера «Как заканчивается Просвещение» (How the Enlightenment Ends), которая, мне кажется, подтверждает мой анализ. Такие же чувства вызывает у меня и книга Киссинджера на эту тему, написанная в соавторстве с Эриком Шмидтом и Дэниелом Хаттенлокером, специалистом по ИИ из Массачусетского технологического института. Его позиция, высказанная в этой работе, представляется мне совершенно ясной.

Об автономии субъекта:

> По мере того как люди все меньше используют свой мозг и больше – машины, они могут утратить собственные способности. Наши навыки критического мышления, письма и дизайна могут атрофироваться.

О знаниях:

> Каким образом обучающаяся машина хранит свои знания, перерабатывает их и извлекает, по-прежнему остается неизвестным. <…> Наука эпохи Просвещения накапливала определенности; новый ИИ аккумулирует двусмысленности.

Таким образом, ChatGPT поднимает вопрос о просвеще-

нии. Как мы видим человека? Как автономного субъекта, способного выявить причинно-следственную связь? Или как индивида, порабощенного игрой корреляций?

В технологиях нет ничего неизбежного. Это инструмент, который должен быть адаптирован к нашему представлению об истине, к нашим политическим и социальным принципам. Но не наоборот…

Прогресс не равен инновации. Прогресс заключается не в слепом принятии всего нового, а в том, чтобы учиться на сделанных ошибках.

Именно это предлагает моя книга, и я надеюсь, что сегодня она актуальна как никогда.

Гаспар Кёниг
апрель 2023

Homo deus

Sorry, I can't help you. «Извините, ничем не могу вам помочь», – вот так лаконично ответил на мою просьбу об интервью Элиезер Юдковский, один из передовых исследователей искусственного интеллекта в Кремниевой долине. Едва ли можно лучше показать пропасть, отделяющую властителей технологий от широкой публики; тех, кто создает алгоритмы, – от тех, кто живет под их властью; и, наконец, тех, кто пишет строки кода, – от тех, кто пытается их понять.

Sorry, I can't help you. Иначе говоря: я хотел бы, но правда не могу. Как варвар, который никогда не занимался программированием, профан, который с трудом справляется с PowerPoint, может понять все тонкости «глубокого обучения»? Цель здесь одна: «Мы делаем мир лучше», *make the world better*. Мы не задаемся вопросами, мы решаем проблемы. Философская болтовня, роскошь праздного рассудка, – не наша забота. Так Горгий ответил Сократу: все это ребячество надо немедленно прекратить.

Sorry, I can't help you. Не удивлюсь, если это сообщение было автоматически сгенерировано Gmail: Юдковский, должно быть, получает кучу подобных запросов. Алгоритм должен научиться распознавать их и выдавать подходящий ответ. Что может быть естественнее искусственного интеллекта, отвечающего на вопросы об искусственном интеллек-

те?

Однако сегодня нам жизненно необходимо говорить друг с другом и помогать друг другу, чтобы понять те грандиозные сдвиги, которые наводят страх на наши сообщества, смешивают карты наших экономических систем, сотрясают наши политические структуры и вторгаются в саму нашу жизнь, вынуждая нас разрываться между надеждой на прогресс и страхом будущего. В прежние периоды технологических перемен мыслители, изобретатели, ученые, инвесторы и политики сосредотачивались в определенных местах, нервных центрах «миров-экономик», описанных Фернаном Броделем. Эти города-миры представлялись не только средоточиями экономики, но также местами высокой культуры. «Блеск, богатство, радость жизни соединяются в центре мира-экономики, в его сердце. Именно здесь, под солнцем истории, жизнь обретает свои самые яркие цвета»[1]. Так все и было: Спиноза посеял семена философии имманентности в Амстердаме, столице золотого века; Адам Смит разработал теорию капитализма в Эдинбурге, в сердце промышленной революции; а Маркс осмыслил классовую борьбу в викторианском Лондоне. Города служили полюсами притяжения и интеллектуальными тиглями: кипящие умы производили в них странные, рискованные и порой чудесные смеси. Сегодня «мозги» разбросаны гораздо шире: никто не может ска-

[1] *Бродель Ф.* Динамика капитализма. Смоленск: Полиграмма, 1993. С. 95. – *Здесь и далее примечания автора, если не указано иное.*

зать, где находится мировой центр искусственного интеллекта. Даже Кремниевая долина с ее спальными пригородами не может сойти за образец «счастливой жизни»; Сан-Франциско, ставший одним из самых дорогих городов в мире, выдавливает молодых изобретателей; калифорнийская контркультура 1970-х вырождается в явное безразличие ко всему, что касается гуманитарных наук, словно бы страсть к переменам уничтожила всякую антропологическую рефлексию, а человек стал просто тестом, из которого можно вылепить что угодно, не считаясь с биологией или историей. Впрочем, несколько лет назад журнал *The Economist* предлагал лидерам «технологий» учиться философии, но безуспешно[2]. Может быть, больше нет никакого города-мира?

Именно для того, чтобы воссоздать что-то вроде такого виртуального города-мира, попробовать навести мосты между буйством технологий и постоянством метафизики, я и предпринял это долгое путешествие в страну искусственного интеллекта. За несколько месяцев я взял интервью у 125 специалистов, оказавшихся отзывчивее Элиезера Юдковского или просто уставших от моей назойливости: исследователей, предпринимателей, инвесторов, преподавателей, чиновников, художников… Я хотел встретиться с ними в их естественной среде – там, где они живут и работают, среди их компьютеров и дорожных пробок, – поэтому и от-

[2] Philosopher kings // The Economist. 2014. Oct 4. URL: https://www.economist.com/business/2014/10/04/philosopher-kings

правился в путешествие вокруг света, двигаясь в западном направлении: Кембридж, Оксфорд, Бостон, Нью-Йорк, Вашингтон, Сан-Франциско, Лос-Анджелес, Шанхай, Пекин, Тель-Авив, Копенгаген и, наконец, Париж. Перемещаясь в эту сторону, то есть по движению солнца, я искренне верил в то, что, потихоньку обкрадывая его, смогу в итоге удлинить свою жизнь на один день, но потом, пролетая над Беринговым проливом, заметил, что линия перемены дат, проходящая по 180-му меридиану, все забрала у меня обратно. Мне удалось повторить ошибку Филеаса Фогга, но в обратном направлении, а это уже объясняется тем, что я все-таки не ученый.

Дорога оказалась неспокойной. Она началась в лаборатории искусственного интеллекта Facebook[3] в Европе, где занимаются фундаментальными исследованиями. Увидев молодую женщину, погруженную в тысячи строк кода, которые выводились на полудюжине экранов, я осознал, что легкомысленно зашел на территорию сакрального и что нельзя попасть в святилище, не подвергаясь рискам и опасностям. Эта исследовательница стремилась найти способ автоматически предсказывать перемещение объектов на улице на основе простого изображения: тронется ли с места эта машина? Начнет ли переходить улицу этот пешеход? Уронит ли этот ребенок мяч? Я инстинктивно отшатнулся, испугавшись, возможно, того, что увижу строчку кода, описываю-

[3] Принадлежит компании Meta, признанной экстремистской в РФ. – *Прим. ред.*

щую мое собственное поведение, словно бы миллионы миллионов нулей и единиц могли охватить всю реальность – и прошлую, и будущую.

Но на этом мои страхи не кончились. Спустя несколько дней Аврелия Жан, молодая специалистка по информатике, окончившая Массачусетский технологический институт (MIT), смело воспользовалась нашим путешествием на высокоскоростном поезде, чтобы познакомить меня с «Питоном», одним из самых известных языков программирования. Однако для меня травмой стало уже то, что на экране компьютера Аврелии я увидел не папки с файлами, а черное окно, заполненное каббалистическими знаками. Дело в том, что она, как и многие ее коллеги, не снисходит до того, чтобы просто кликать мышкой по слишком удобным иконкам пользовательского интерфейса. Аврелия работает под капотом машины, поближе к ее первичным функциям. Она дает ей инструкции в форме кода. Например, вместо того чтобы открывать папки, когда надо найти документ с текстом, она приказывает компьютеру, на понятном ему языке, его отыскать. Ей кажется, что это более естественный способ общения с информационным инструментом. Мы же, профаны, подобны детям, которые, если надо выполнить какие-то арифметические операции, вынуждены прибавлять и вычитать куски пирога, – нам нужна определенная репрезентация (кстати, именно она определила успех компаний Microsoft и Apple в 1980-х). Аврелия же манипулирует непосредствен-

но цифрами и обходится без дополнительного символического уровня. «Так быстрее», – говорит она мне, барабаня по клавиатуре.

ИИ умножает число проблем, которые отпугивают исследователей. Похоже, он развивается в самых густонаселенных городах: простояв десятки часов в пробках, я понял страсть гиков к «умному городу» и беспилотным автомобилям. Главное же, что самая революционная технология последних десятилетий привязана к довольно таинственной науке. В этой области очень мало общих работ, ориентированных на неофитов. Не буду делать вид, что преодолел хотя бы введение к библии специалистов по компьютерным наукам – «Искусственному интеллекту» Стюарта Рассела и Питера Норвига: после горстки определений и исторической справки изложение вскоре стало слишком техническим, неприятно напомнив мне обо всех причинах, заставивших меня уйти из последнего научного класса колледжа, где я промучился несколько недель, и перейти в литературный. Тем не менее благодаря чтению и разговорам на тему ИИ я, кажется, приобрел своего рода «окраску», если пользоваться выражением Монтеня, которым он обозначал наши познания, всегда остающиеся несовершенными. Окраску, необходимую, хотя и недостаточную, чтобы делать вид, что философствуешь. Окраску, на которой оставили свой след случайность, локальные открытия, наваждения: в репортаже нужно смириться с определенной долей удачи и

неудачи, откровения и невежества. Мой ежедневник, когда я приезжал в какой-либо город, часто заполнялся лишь по мере встреч. В своем исследовании я придерживался той «серендипности»[4], которую ИИ как раз хотел бы уничтожить.

Я провел четыре недели на Западном побережье, и только однажды мне довелось зайти в кабинет, полки которого прогибались под классикой: в фонде Питера Тиля в Лос-Анджелесе я внезапно оказался в своей тарелке – среди томов Сен-Симона в издании «Плеяды» и работ Рене Жирара. В приемной я с изумлением обнаружил экземпляр «Речи о положении великих» Паскаля. Возможно, предпринимателю в сфере технологий, чрезвычайно успешно привлекающему средства, нелишне время от времени вспоминать о различии между «величием по установлению» и «естественным величием». Первое, связанное с социальным статусом, требует вполне обоснованного почитания могущественных людей, однако не может определять реальных человеческих качеств, которые связаны со вторым. Паскаль – не революционер: он не предлагает низвергнуть великих мира сего, но призывает нас держаться «двоемыслия», которое умеет отличать социальные условности от моральных достоинств. К этой рекомендации полезно прислушаться нашим предпринимателям, которые, одевшись в наряды ложной аутентичности, сотканные из эмодзи и селфи, делают вид, будто игно-

[4] Серендипность (от *англ.* serendipity) – интуитивная прозорливость, способность делать выводы из случайных наблюдений. – *Прим. ред.*

рируют отношения власти и капитала, управляющие их отношением к другим. Может быть, они тоже путают величие по установлению с естественным величием?

Дело в том, что этикет Кремниевой долины ни в чем не уступает этикету былых королевских дворов. *Cool*, «крутизна», породила свои собственные кодексы, соблюдать которые жизненно необходимо, если хочешь преуспеть или просто выжить в этой экосистеме непрестанной конкуренции. Я очень скоро понял, что электронное письмо, каким бы горячим и откровенным оно ни было («Привет, Марк! Я французский философ»), обречено на вечное молчание. Как всегда, не нужно верить рекламе: нет там ничего горизонтального, текучего или прозрачного. *Think different*, «думай иначе», но не переборщи. Встречи можно добиться только после долгих переговоров через общих знакомых; сначала надо попросить, чтобы тебя представили, – это необходимое условие, которое само требует хитрых риторических уловок. Если ты забыл восклицательный знак или смайлик, это уже может говорить о непростительной нехватке энтузиазма. Нужно быть серьезным, казаться игривым, доказывать свой пыл и намекать на преданность, и все это одновременно и в трех пунктах. Сам Сен-Симон покажется на этом фоне простаком! В мире гугл-календаря никуда не делись армии секретарей, которые отфильтровывают посетителей и просителей. Личная беседа – вот что работает в эпоху «пост» лучше всего. Кстати, один француз, давно уехавший в Сан-Франсис-

ко, написал точную и забавную статью о строгих правилах «Кремниевого Версаля», начиная с тайм-менеджмента и заканчивая формулами вежливости[5].

Достаточно зайти в ресторан Madera в Rosewood Sand Hill, центре венчурного капитала, – в «храм сделок», как съязвил пригласивший меня инвестор, – чтобы заметить, насколько кодифицированным и иерархическим остается мир Кремниевой долины. Прежде всего, это место невозможно найти, оно не определяется по GPS. «Это сделано специально, – сказал мне мой провожатый, впиваясь зубами в самый дорогой на планете гамбургер. – Это мир инсайдеров. Здесь у нас не демократия». Через панорамные окна открывается спокойный вид на заповедник Jasper Ridge; под голубым небом калифорнийского лета расположились островки леса, колышущиеся бархатистыми волнами. Можно разглядеть секвойи с красноватыми стволами и мускулистыми ветками, качающимися на ветру. На переднем плане под солнцем блестят высаженные в безупречном порядке оливковые деревья. Несколько десятков скромных домов выстроились вдоль тенистых аллей. «Это самые крупные в мире фонды венчурного капитала: Sequoia Capital, Menlo Ventures, Schlumberger, Makena Capital, Andreessen Horowitz, Coaetue Management, Silver Lake Partners, Kleiner Perkins… Деньги прямо здесь, вокруг нас. Десятки миллиардов, ежесекундно готовых к инвестированию». Деньги стали экологичными, они обманы-

[5] https://medium.com/@romainserman/silicon-valley-etiquette6934cf6f8f73

вают тех, кто все еще ищет их на верхотуре стеклянных небоскребов. Предприниматели совершают свое паломничество пешком, от двери к двери, как дети, просящие сладости на Хеллоуин. Их легко узнать: они потрудились надеть пиджак, они быстро говорят и слишком много улыбаются. Тогда как инвесторы, у которых в руках ключи к амбициям предпринимателей, принимают их расслабленно, чаще всего они в кроссовках кислотных цветов. Это капитализм в джинсах, какая-то энная версия черных сюртуков Фуггеров или безупречных костюмов Ротшильдов. Вопреки тому, что повторяют нам многоречивые гении разрушения, выступающие на TED Talks, мир не так уж и меняется.

Паскаль приходит к выводу: «Я не обязан уважать вас за то, что вы герцог, но я должен снять перед вами шляпу». То же самое в Кремниевой долине: если вы венчурный капиталист, это не значит, что я обязан вас любить, но я должен вас *like*.

Последний урок, который преподал мне мой гид, прежде чем сбежать на какую-то встречу (разумеется, срочную), заключался в том, что ни предприниматели, ни инвесторы не имеют ни малейшего представления о социальном и политическом влиянии создаваемых ими технологий. С его точки зрения, искусственный интеллект сопряжен с изрядной дозой «поверхностного интеллекта». Чтобы сохранить первичную связь между технологической инновацией и философской рефлексией, нужно, чтобы в Rosewood Sand Hill было

больше Паскалей. И разве сам Паскаль не был предпринимателем, создателем первых городских автобусов, «карет с пятью этажами»?

Но в то же время нужно сопротивляться навязчивому искушению, которое ведет к технофобии. Когда мы смотрим в прошлое, пророки апокалипсиса всегда кажутся нам смешными: так, Поль Валери почти столетие назад разоблачал «коварную отраву» технического прогресса и (уже тогда!) сетовал на исчезновение свободного времени, склонность просматривать книги, а не читать их, диктатуру эмоций… «Ни почта, ни телефон не докучали Платону»[6], – сожалеет Валери. Бедный поэт, которого прервал почтальон! Что бы он подумал об уведомлениях сетевых сервисов и твитах? Если хочешь попытаться понять свою эпоху, не оболгав ее, надо заставить и себя, и читателя совершить над собой усилие.

Тем более что искусственный интеллект должен быть мечтой всякого философа. Разве не удобно было бы создать мыслящую машину, которая избавила бы нас от логических ошибок, индивидуальных предрассудков, концептуальных заблуждений? Создать алгоритм, который посчитает истину и даст нам наконец, после тысячелетий однообразных споров, ответ на наши самые важные вопросы о смысле жизни? Концептуальное мышление – это, по сути, всего лишь приближение; тогда как полная система символов, управляемая научными законами, позволяет максимально подойти к

[6] *Valéry P.* Le Bilan de l'intelligence [1935]. Allia, 2016.

истине. Первым об этом начал мечтать Лейбниц, гений математики и метафизики, который искал формулу машины для подсчета мыслей, «универсальной характеристики», ведущей к правильному рассуждению. Гигантская комбинаторная машина, которую Лейбниц назвал *calculus ratiocinator*, автоматически прогнала бы все химеры разума. В этом совершенно рациональном мире «больше не было бы нужды в дискуссии двух философов более долгой, нежели дискуссия двух математиков, поскольку им достаточно будет взять в руки перо, усесться за свои счетные доски и сказать друг другу: подсчитаем!»[7]

Такой же идеал мы встречаем у всех великих предшественников искусственного интеллекта: Гильберта, Фреге и, конечно, Алана Тьюринга – все они поддерживали тесные отношения с логикой и аналитической философией[8]. Если логическая точность – то, что позволит человеческому разуму лучше воспринимать реальность, тогда превращение реальности в цифровую комбинацию могло бы успешно заменить собой человеческий разум. Отсюда насмешливый ответ Deep Thought, суперкомпьютера из романа «Автостопом по Галактике»: когда его попросили решить «последний вопрос жизни, Вселенной и всего остального», он сказал: «42». Это число стало потом легендарным, оно продолжает то и де-

---

[7] *Leibniz W.* Nova methodus pro maximis et minimis. 1668.

[8] Интеллектуальную историю ИИ см. в: *Davis M.* The Universal Computer. W.W. Norton, 2000.

ло всплывать в разговорах гиков. Разве не наступил бы покой, если бы все смыслы, пронизывающие наши жизни, удалось свести к ничего не значащей цифре?

Это родство информатики и логики сохраняется и сегодня, навязчиво преследуя посетителя причудливого здания департамента компьютерных наук в MIT: войдя в кабинет Лесли Келблинг, ветерана исследований в области ИИ, я обнаружил ее в окружении книг по аналитической философии – она увлекается Уиллардом Куайном. Этот подход весьма далек от интеллектуальной традиции континентальной Европы. В конце прошлого века Делёз довел это противопоставление до предела, определив философию в качестве «производства концептов», то есть деятельности, связанной скорее с творческим процессом, чем с научной строгостью. Концепт невозможно свести к последовательности единиц и нулей: он проецирует на наш мир новый смысл, дополнительную перспективу. Такой благодатный момент часто наступает, когда, заново прорабатывая какой-нибудь сухой трактат по метафизике, мы внезапно «понимаем» его смысл. Все концепты тогда естественным образом складываются воедино, и глаз начинает скользить по странице с неожиданной легкостью. Словно бы аргументы, разрабатываемые на протяжении многих глав, были не столько этапами одного логического пути, сколько линзами, постепенно меняющими наше восприятие вещей. Мы напрягаемся изо всех сил и рвем волосы на голове, стремясь понять положения «Этики»

Спинозы, сражаемся с ними, пытаясь проанализировать их цепочку, и вдруг без фанфар и предупреждений наступает такой момент, когда тезис *Deus sive Natura*, «Бог или Природа», предстает перед нами во всей своей концептуальной глубине. Тогда можно перечитать положения, которые казались столь загадочными, – и они окажутся на удивление очевидными. В этом феномене задействуется не только разум, но и – каким-то непонятным образом – все наше тело. «Модификация его отношений движения и покоя», – сказал бы Делёз.

И наоборот, мне вспоминается замешательство, которое я испытал, когда проходил курс философии в Колумбийском университете, где меня просили разбить фразы на уравнения, отыскивая «истину» высказывания в игре предпосылок и выводов: как эти школьные упражнения могут иметь хоть какое-то отношение к мышлению? С другой стороны, американцы на то, что они сами назвали «французской теорией», смотрят с изрядным скептицизмом, поскольку часто видят в ней лишь поэтическое фанфаронство. Вполне возможно, что неприятие европейцами аналитической философии в какой-то мере позволяет объяснить их инстинктивное недоверие к ИИ. Так или иначе, следует помнить о том, что ИИ – не только промышленная технология, но и прежде всего философский проект, нацеленный на понимание мира.

Но для того чтобы отправиться в это долгое путешествие, у меня были причины и более личного свойства. Я либерал

и защищаю идею автономного индивида, свободного в своих решениях и ответственного за свои действия, то есть такого индивида, который должен применять свободу воли в той или иной ее форме. В основании наших обществ еще с эпохи Просвещения лежит эта идея, оправдывающая несколько вещей сразу: индивидуальные права, рыночные механизмы, право голоса и уголовную юстицию. Верховный суд США нисколько не ошибся, когда увековечил свободу воли в качестве самого́ условия правовой системы, позволяющего человеку делать выбор между добром и злом, желаемым и порицаемым, дозволенным и запрещенным.

Но сегодня это разумное здание трещит по швам. В своем бестселлере «Homo deus» историк Юваль Харари выдвигает головокружительный прогноз: промышленное приложение искусственного интеллекта ускорит и конкретизирует исчезновение свободы воли, ставшее предметом современных наук. «В начале третьего тысячелетия либерализму угрожает не философская идея, отрицающая существование свободных индивидуумов, а конкретные технологии. В самом скором времени нас ожидает нашествие чрезвычайно полезных устройств, приспособлений и структур, которые не оставят места свободной воле индивидов. Выживут ли в этих условиях демократия, свободный рынок и права человека?» Харари, который требует полностью переопределить наши идеологии и институты, считает, что нет, не выживут. Таким образом, контролируя наше поведение и руководя самыми тай-

ными нашими помыслами, ИИ сможет подорвать либеральное основание наших обществ, разрушив само понятие индивидуальности. Если алгоритм знает меня лучше, чем я сам, и предлагает мне более рациональные решения, чем я мог бы принять самостоятельно, если мириады объединенных в сеть объектов делают мою способность к решению излишней, предлагая мне жизнь комфортную и предопределенную, если я постепенно перестаю быть движущей силой собственных действий, зачем мне понадобится право голоса и буду ли я подлежать хотя бы малейшей уголовной ответственности? ИИ может добить свободу воли, а вместе с ней и кантовский идеал автономии субъекта. Триумф благополучия означал бы тогда отречение от свободы – свободы выбора, свободы восстания, свободы ошибаться, «свободы заблуждаться», о которой говорил Джон Стюарт Милль[9].

Этот вывод, поспешный и радикальный, стал для меня как отправной точкой, так и пунктом назначения, поскольку мои путешествия окончились дискуссией с Харари в его святилище в Тель-Авиве. Не рискует ли либерализм, прославляя чудеса технологий, потерять самого себя? Мне кажется, что апологеты индивидуальной свободы настолько устали от векового сражения с луддитами самых разных мастей, настолько ослеплены своей страстью к инновациям, что просто отказываются видеть ту фундаментальную опасность, кото-

[9] Эту изящную формулу – «the right to err» – ввел Исайя Берлин, когда комментировал работы Дж. С. Милля.

рую ИИ представляет для самого понятия индивида. Заметным исключением выступает лишь Питер Тиль, знаменитый предприниматель и открытый либертарианец, который охотно заявляет, что «ИИ – штука коммунистическая», поскольку требует централизации и нормативов. Если ИИ предвосхищает, регулирует мое поведение, манипулирует им и даже моими самыми тайными помыслами, если он способен, как воображает Юваль Харари, «хакнуть людей», нельзя ограничиться классическим аргументом о свободном рынке. Как потребитель может оставаться прав, если сами основания его решений сфабрикованы алгоритмом? Как мы можем быть ответственными взрослыми людьми, если самые главные наши решения определяются нашей включенностью в сеть? Какая разница между Google и Коммунистической партией Китая, если они используют одни и те же техники *nudge*, преследуя одни и те же утилитаристские цели?

За свою неспокойную жизнь либерализм пережил множество кризисов. Оправившись от краха 1929 года, он был вынужден отказаться от слишком радикальной позиции *laissez-faire* и допустить необходимость регулирования [10]. Сегодня же ИИ должен подвести нас к вопросу о примате индивидуальной рациональности: возможно, не всякий добровольный (то есть свободный от принуждения) выбор является «сво-

[10] Таковы были плоды коллоквиума Липпмана 1938 года, который в значительной степени определил неолиберальное мышление послевоенного периода. Об этом историческом моменте см. работу: *Audier S.* Le Colloque Lippmann. Aux origines du «néolibéralisme». Le Bord de l'eau, 2008.

бодным». Если не изучить этот вопрос, рухнуть может все здание либерализма, построенное за три столетия.

Грозит ли нашим свободам запрограммированное моральное устаревание?

1
Механический турок

Почему ИИ – это иллюзия

«Это просто магия…» – вот все, что мне удалось сказать, когда основатель одного израильского стартапа показал мне свою программу аудиораспознавания эмоций. Мы сидим не в гараже в Кремниевой долине, здесь нет настольного футбола и колы без сахара по первому требованию. Мы где-то в Тель-Авиве, между автотрассой и недостроенным зданием. Вход найти нелегко: он прячется за магазином дешевой бытовой техники. Комната для переговоров пустая и пыльная, на столе пластиковые стаканчики – все это больше похоже на контору по экспорту-импорту, чем на современную инновационную компанию. Однако же Юваль Мор, основатель стартапа, только что показал мне, как его алгоритмы могут воспринимать всю палитру чувств, содержащихся в той или иной фразе, тональности, даже в шепоте. Пруст больше не нужен: приложение в реальном времени выдает все нюансы разочарования, одиночества или тайного удовольствия. Более того, недавний эксперимент позволил установить корреляцию между тембром голоса и симптомами сер-

дечной недостаточности. На основе нескольких звуков ИИ может проинформировать вас как о любовных тайнах, так и о рисках остановки сердца. А поскольку нужно еще и деньги зарабатывать, он может рассказать специалистам по телемаркетингу о настроении их клиентов. А что, если вскоре появится сервис для выявления лицемерия? Жить в обществе станет куда сложнее…

Трудно не восхищаться чудесами ИИ. Ежедневно нам объявляют о том, что роботы превзошли самых опытных врачей в диагностике рака или же что был автоматизирован труд журналистов. Исследовательские институты выпускают один доклад за другим, в более или менее катастрофическом тоне рассказывая нам о том, что нет такой области деятельности, которую пощадит разгул технологии[11]. Ярлык «ИИ» стал волшебным словом, паролем, позволяющим продать любую идею инвесторам, которые не поспевают за скоростью этих преобразований. Мне недавно рекомендовали упомянуть ИИ в научном труде по философии, чтобы повысить свои шансы на успех. Что же касается настоящих специалистов, знатоков *deep learning* и нейронных сетей, то компании дерутся за них, а их годовые зарплаты нередко переваливают за миллион долларов. В Китае муниципалитеты обещают золотые горы молодым инженерам-программистам. Два-

[11] Например, в 2017 году Центр исследований экономической политики (Center for Economic Policy Research) в Лондоне определил процент внедрения ИИ в разных секторах экономики – от 42 % в телекоммуникациях до 18 % в туризме.

дцать лет назад царский путь к богатству требовал способности придумывать экзотические финансовые продукты; сегодня же нужно уметь писать код.

В США ИИ перестал быть особой темой, привлекательной или отталкивающей, поскольку он полностью интегрирован в повседневную жизнь: о нем говорят так же, как об электричестве или интернете. В «мозговых центрах», которые я посетил в Вашингтоне, ИИ не составляет предмета отдельного исследования, он проник в самые разные области, от экономики до политики и военного дела. На автотрассе, связывающей Кремниевую долину с Сан-Франциско, можно увидеть рекламные щиты, расхваливающие ИИ-компании так, словно речь идет о последней модели барбекю: «Brighterion: искусственный интеллект для критических задач»; «Darktrace: мировой лидер в кибер-ИИ». Мода на ИИ захватила всех: Шахид, водитель Uber, который вез меня ранним утром в Беркли в офис Стюарта Рассела, знаменитого специалиста по информатике, поведал мне, что сам ходит на курсы программирования, поскольку «ИИ – это будущее». С ИИ знакомы дети: в первом эпизоде мультфильма «Суперсемейка» герой вынужден сражаться с самообучающейся боевой машиной. И даже рестораны смотрят в эту сторону: в заведении Situ в Сан-Франциско мне довелось отведать суп из карамелизированной моркови по рецепту Натана Мирвольда, одного из бывших руководителей Microsoft, который теперь применяет науку о данных в кулинарии.

Конечно, есть интеллектуалы, которые борются с технологическим капитализмом, эдакие пережитки калифорнийских 1960-х. Их труды можно найти в знаменитой независимой библиотеке Сан-Франциско City Lights, где царит изысканная атмосфера легкого запустения. Вот, к примеру, книги, которые стоят на самом первом стеллаже, в тесном строю, словно бойцы старой гвардии: «Интернет как оружие. Что скрывают Google, Tor и ЦРУ»[12], «Новые Темные века: технология и конец будущего», «Habeas Data: распространение технологий надзора», «Тюремщики интернета» и т. д. Иными словами, рассуждения, часто довольно однообразные, о надзоре – понятии, позаимствованном из трудов французского философа Фуко, – никуда не делись… Однако в целом американское общество, похоже, не разделяет опасений прогрессистской элиты. В Музее де Янга, стоящем посреди парка «Золотые ворота» в Сан-Франциско, в экспозиции, посвященной «культу машины», картины прецизионистов 1930-х, прославлявших паровую машину, выставили рядом с цитатами из современных трансгуманистов. Газетные статьи межвоенных лет напоминают о том, что страх машины существовал всегда: «Г-н Робот превзойдет своего хозяина», – вот о каких ужасах рассказывали тогда в передовицах. Идея выставки оказалась как нельзя более прозрачной: нынешние страхи цифровизации сравнимы с прежней бояз-

[12] *Левин Я.* Интернет как оружие. Что скрывают Google, Tor и ЦРУ. М.: Individuum, 2019.

нью механизации, которая сегодня кажется нам смешной. После осмотра экспозиции посетителям Музея де Янга предлагали выбрать для описания технологии три слова из тридцати. Вот что пришло им в голову: креативность, революционность, эффективность, прогресс. А вот слова, которые остались неиспользованными: надзор, загрязнение, неравенство, отчуждение. Похоже, что ИИ следует естественным путем неумолимой инновации. Завтра, опутав всю нашу жизнь тысячами подключенных друг к другу объектов, он станет привычным и невидимым.

Итак, общество сначала изумляется, возмущается, впадает в панику, бунтует, а потом привыкает и теряет интерес. Выступления на TED, в которых пророки нового поколения объясняют нам, почему мы стали бесполезны или как нам будет скучно, когда мы сможем жить по сто лет, даже не развлекают нас: все это уже было. Общество вскоре полностью «переварит» ИИ.

Мы свыклись с магией. Нужно быть въедливым занудой, чтобы все еще стремиться разгадать этот фокус.

А это как раз мой случай…

От барона фон Кемпелена до компании Amazon

В этом тумане искусственного интеллекта, окутавшем весь мир, компания Amazon подала нам ценный сигнал, назвав свою платформу микрозадач Amazon Mechanical Turk – «Механический турок Amazon». Сотни тысяч внештатных работников, называемых «турками», получают вознаграждение за выполнение в интернете простейших задач (например, за сортировку изображений), результаты которых поступают в исследовательские или производственные системы ИИ. Почему в наше время, когда малейший культуралистский намек жестко подавляется, было выбрано это странное название – «Механический турок»?

Дело в том, что в 1769 году именно так венгерский изобретатель Вольфганг фон Кемпелен назвал свой шахматный автомат – марионетку, одетую в турецкий костюм. Этот механический турок сумел поставить шах и мат многим известным шахматистам того времени, а также некоторым историческим личностям, в частности Марии Терезии, Наполеону Бонапарту и Бенджамину Франклину. Механический турок, сидевший за своим внушительным ящиком-столом с шахматной доской, перемещал фигуры резкими движениями и мог даже проявлять во время партии определенные эмоции, например таращить глаза, качать головой или ше-

велить пальцами. Блестящий тюрбан, суховатые черты лица, длинные османские усы – все это дополняло драматическое напряжение. Механический турок прославился по всей Европе; впоследствии он достался Иоганну Мельцелю (изобретателю метронома), который уехал с ним сначала в Лондон, а потом в США. На заре промышленной эпохи, когда в обществе вдруг возникла повальная мода на автоматы, а математик Чарльз Бэббидж только-только представил свои революционные счетные машины с перфокартами, люди спрашивали себя, не изобрел ли фон Кемпелен механическое мышление. Если человек – это просто машина, как утверждали Ламетри и многие другие философы Просвещения, почему машина не может стать человеком? Вопрос о «сингулярности», который мучает нас сегодня, далеко не нов. Еще два столетия назад он привлекал к себе неподдельное внимание. Механический турок, с точки зрения современников, был первым искусственным интеллектом.

Конечно, дело в трюке, причем весьма простом. Внутренняя часть ящика перед представлением всегда открывалась, и зритель видел там лишь сложный механизм из шестеренок и приводов. Однако ловкая игра зеркал и двойное дно позволяли скрыть профессионального шахматиста из плоти и крови, который, сидя в ящике, выполнял сложные движения, передвигая фигуры. То есть первый ИИ был грубым обманом, и сегодня можно только удивляться, почему он почти целый век пользовался таким оглушительным успе-

хом. Amazon вдохновился этой историей, чтобы остроумно напомнить нам о том, что за магией алгоритмов скрывается значительный человеческий труд, позволяющий собирать, обрабатывать и извлекать данные. Возможно, аудиораспознавание эмоций однажды покажется нам такой же грубой уловкой, как и трюк фон Кемпелена. Неужели мы относимся к новым технологиям с той же наивностью, что и светские дамы XVIII века, которые млели перед деревянным автоматом? Гуманоиды выступают сегодня на конференциях, а робот София получил гражданство в Саудовской Аравии, но действительно ли они намного совершеннее своего общего предка – механического турка?

Мне захотелось разобраться в этом вопросе. Исходный шахматный автомат сгорел в 1854 году во время пожара в музее Филадельфии, но существует его точная копия, которую крайне редко показывают публике. Я отправился ее исследовать. Не мог же я изучать вселенную ИИ, не пожав руку механическому турку?

Северный пригород Лос-Анджелеса, между Адамс-хилл и Гриффит-парком. Район представляет собой редкое сочетание промышленных зон, одноэтажных домов в самых разных стилях и магазинов органических продуктов; повсюду растут пальмы, а вдали возвышаются пустынные холмы. Трудно представить что-либо более американское: каждый в этом открытом пространстве создает свое маленькое царство. Демократия увенчала короной голову каждого гражда-

нина. Я вступаю в одно из таких княжеств – обширный ангар, в котором рабочие что-то делают под визг электропил. Меня встречают картонные роботы: Пьеро, сидящие на Луне, головы Микки-Мауса, светящиеся рекламы выступлений фокусников и расположенные в шахматном порядке зеркала – в них отражается мое помятое лицо. Я пробираюсь в маленькую комнату, где неожиданно царит полная тишина. Это кабинет настоящего антиквара, элегантный и обшитый деревом. Здесь кучи бильбоке, черепов, коробок с игральными костями, кожаных чемоданов, вееров, подзорных труб и карточных колод. Но вот настоящее сокровище – целая толпа автоматов в натуральную величину: Гудини дает автограф своей гипсовой рукой, Вильгельм Телль потрясает луком, павлин протягивает мне в клюве даму пик. В кресле красного бархата посреди всех этих созданий восседает иллюзионист Джон Гуган, который уже многие годы воспроизводит химеры прошлого и изобретает химеры будущего. Рядом с ним – главный экземпляр: турок, абсолютно бесстрастный в своей белой меховой шубе, готовый начать игру.

Вот уже сорок лет, как Джон Гуган пытается вернуть жизнь турку, обшаривая библиотеки Берлина, Парижа и Лондона, чтобы среди сотен книг того времени найти сведения, позволяющие восстановить исходный механизм. Это стало делом всей его жизни, о котором он периодически рассказывает на конференциях специалистов по компьютерным наукам. Джон, с его глухим голосом и запавшими глазами

под густыми бровями, настолько похож на волшебника, что так и хочется спросить: не сделан ли он сам из шестеренок, приводных ремней и силикона?

Правда ли, что внутри ящика находился человек? А как он мог поместиться в таком узком пространстве? Джон следует этическому кодексу своей профессии: он отказывается рассказывать о том, как действует иллюзия. Он сомневается даже в том, стоит ли доверять свои открытия тексту; возможно, секрет турка исчезнет вместе с ним, разве что потом какой-нибудь его последователь будет исследовать Джона Гугана так же, как он сам изучал Вольфганга фон Кемпелена… Я стою перед турком, трогаю его, открываю дверцы ящика, но он все равно остается для меня непостижимым. Однако нужно как-то сохранить ту ничтожно малую иррациональную долю сомнения, допускающего, что машина и впрямь могла думать… Ведь парадокс в том, что именно это сомнение, эта потребность в магии и составляют нашу человечность.

«В моем ремесле, – рассказывает Джон, – меня поражает то, насколько первобытным остается человеческий разум». Как легко обмануть публику, отвлекая ее внимание довольно простыми сигналами. На самом деле сегодня это еще проще: в нашей цифровой среде полно постоянных отвлекающих факторов, все больше и больше снижающих способность к концентрации. «Но это не относится к детям», – уточняет Джон. Они менее чувствительны к кодексам нашего повсе-

дневного взаимодействия и еще не до конца прошли социальную дрессуру, поэтому не так доверчивы Фокусник показывает голубя, который взлетает с правой руки, но ребенок продолжает смотреть на левую, не обращая внимания на спектакль, а следуя лишь собственному размышлению. Чем старше ребенок, тем проще им манипулировать.

И что же тогда можно сказать об ИИ? «Это иллюзия, то есть моя вселенная». Полезная иллюзия. Да и, по словам Джона, разве турок не дал импульс промышленной революции? Машина порождает машину, магия питает прогресс.

Покидая кабинет Джона Гугана, я снова погружаюсь в истому калифорнийского лета: одинокий пешеход в городе, созданном исключительно для машин. Теперь я лучше понимаю ставки, сделанные на ИИ. Это иллюзия. Задача не в том, чтобы досконально понять, как он работает, а в том, чтобы вопреки всем бредовым идеям, которые он пробуждает, сохранить холодный рассудок. Рассудок ребенка, который смотрит в другую сторону…

Machine learning, deep learning, reinforcement learning, unstructured learning[13] – все эти термины смешивались в головокружении небоскребов Нью-Йорка, где я начал свои странствия. Кроме того, прибыв на самолете в Бостон, я слишком смело сразу полез купаться и заработал отит, что не упрощало мне задачу понимания ИИ. Беспрестанный гул

[13] Базовые определения этих терминов словаря машинного обучения см. здесь: https://developers.google.com/machine-learning/glossary/

Нью-Йорка заглушала канонада в моем левом ухе; я ходил на встречи, опустошая запасы анальгетиков и пытаясь поворачиваться к собеседникам здоровым ухом. Все, что я понял, свелось к тому, что пресловутый «ИИ» или, по крайней мере, последнее поколение алгоритмов может более или менее автономным образом копаться в массе данных, извлекая из них определенные закономерности и делая прогнозы. Говорите громче, пожалуйста.

Смехотворность моей задачи и абсурдность самого моего положения стали особенно ясны по дороге в офис *IBM Watson* на Астор-Плейс, где на входе меня поджидал огромный кролик Джеффа Кунса. Пытаясь сформулировать вопросы по информатике, которые преследовали меня, но никак не давались, я вдруг заметил, что забыл запонки, а потому рукава моей рубашки болтались как кружевные манжеты версальских маркизов. Мой горячечный разум принялся изо всех сил решать этот важнейший вопрос. У меня нет строгих привычек в одежде, но должен признаться, что обычно я иду против среды, то есть на встречу с предпринимателями из технологических компаний прихожу в костюме, а на встречу с банкирами – в футболке. (Спустя несколько недель в салат-баре в Сан-Франциско основательница одного стартапа похвалила меня за старомодный имидж: «Как же приятно видеть человека в пиджаке…») До встречи в IBM у меня оставалось пять минут. Я заскочил в большой супермаркет, но ничего там не нашел, а потом в прачечную, где хозяйка

придумала гениальный и совершенно нью-йоркский выход – просто зашить рукава. Мы поговорили о Румынии, откуда она родом и где живет семья моей жены. Я тепло обнял ее на прощание, а потом отправился в IBM в более или менее приличном виде, радуясь столь хитрому решению. Пусть подкладка и не в порядке, но форму я, по крайней мере, сохранил.

IBM – гигант программного обеспечения, прославившийся тем, что его суперкомпьютер Deep Blue победил Гарри Каспарова. Watson – последний продукт их программы ИИ, способный выиграть в общекультурной телевикторине Jeopardy![14]. Сегодня программисты IBM готовят свою машину к риторическим сражениям с людьми. Причем Watson внедряется и в виде коммерческих продуктов, которые продаются разным компаниям, желающим улучшить обработку данных. Он использует несколько слоев анализа: публичный ИИ, обрабатывающий сетевую информацию (например, из «Википедии»), специфичный для каждой конкретной области ИИ (например, финансов), затем частный и специфичный для каждого клиента ИИ (например, для компании J.P. Morgan). Подобная комбинация позволяет производить все более экспертный и независимый ИИ, способный накапливать и синтезировать знания и опыт, приобретенные в определенной сфере деятельности. Например, Watson управляет переносом знаний о нефтяных платформах, отве-

[14] Напоминает игру «Сто к одному». – *Прим. ред.*

чая на технические вопросы новичков. Вы спрашиваете, каким должен быть максимальный вес вертолета при приземлении? Нет нужды обращаться к более опытным коллегам, вам ответит наш компьютер![15]

Все эти практические примеры, утопающие в многословии коммерческого пиара, не слишком помогли мне понять природу подобных технологических достижений. Но, наконец, появилась мадонна ИИ – Франческа, известная специалистка по компьютерным наукам из Падуанского университета, сегодня она работает в исследовательском подразделении IBM. Может быть, ее объяснения показались мне настолько прозрачными лишь потому, что Франческа – рыжая элегантная итальянка, выгодно отличающаяся своей человечностью в этом мире нердов – ботаников, помешанных на технологиях? Так или иначе, краткий курс, прочитанный мне в конференц-зале IBM, где Франческа по старинке писала своим округлым почерком на белой доске, позволил четко организовать в уме все те загадочные понятия, которые я долго собирал в чтении и обсуждениях. Наконец-то все стало обретать смысл… А потому я поделюсь здесь этим безупречным уроком, который специалистам, возможно, покажется слишком упрощенным, но для меня в моем долгом странствии стал непреложным ориентиром. Кстати, нижеследующие строки покажутся менее сухими, если вы будете читать

[15] *High R*. The Era of Cognitive Systems: An Inside Look at IBM Watson and How It Works. IBM Corporation. Redbooks, 2012.

их с итальянским акцентом.

Вначале было логическое правило. Термин «искусственный интеллект» существует с 1950-х годов[16] и в той или иной степени смешивается с понятием информатики как науки. Цель его проста: создать неорганическую копию человеческого интеллекта. За свою не слишком долгую историю ИИ пережил немало приключений и несколько «зим», когда его считали умершим[17]. Долгое время он мог действовать только по правилам, созданным людьми, то есть по пресловутым алгоритмам, которые всегда не более чем сложные руководства. Всем известный Deep Blue, выигравший в конце концов у Каспарова в шахматы в 1997 году, использовал брутфорс, то есть перебирал миллионы возможных комбинаций за несколько секунд. Такой ИИ представляет ту или иную ситуацию в символьном виде, а затем строит рассуждение, которое может завершиться тем или иным решением. По сути, это способ индустриализации логических умозаключений, идеально подходящий для таких закрытых систем, как шахматы. Сегодня такой ИИ называют GOFAI, *good old-fashioned AI*, «старый добрый ИИ».

В своем минимальном варианте ИИ сводится, таким образом, к сумме наших знаний в области информатики. В мак-

---

[16] Термин был изобретен Джоном Маккарти в 1955 году. В следующем году на знаменитой конференции в Дартмут-колледже были заложены основания ИИ как академической дисциплины.

[17] Краткое введение см. в: *Wooldridge M.* Artificial Intelligence. Penguin, 2018.

симальном – это сам человеческий интеллект, то есть все, что компьютерная программа пока делать не умеет; и наоборот, «как только она начинает работать, это больше не называется ИИ», – объяснял Джон Маккарти. Но между двумя этими крайностями в обыденном языке ИИ стал обозначать вполне определенную технику, а именно *machine learning*, машинное обучение.

Собственно, настоящий прорыв, объясняющий массовое распространение технологий ИИ и популярность этого термина, произошел в самом начале текущего столетия, когда информационные системы приобрели возможность обучаться самостоятельно, не следуя заранее установленным правилам. Эта цель была поставлена с самого начала информатики, однако добиться удовлетворительных результатов не удавалось. Успешное решение этой задачи объясняют три фактора: внезапно возникшее благодаря интернету изобилие данных, стремительное увеличение мощности компьютеров и открытие заново «нейронных сетей», то есть определенного способа конструирования информационных связей, при котором точки обработки данных в значительной мере независимы друг от друга, напоминая этим в какой-то степени нейроны нашего мозга.

Машинное обучение, в свою очередь, подразделяется на несколько техник в соответствии с уровнем вмешательства человека: «обучение с учителем» (*supervised learning*, под контролем программиста), «обучение с подкреплени-

ем» (*reinforcement learning*, когда машина «вознаграждается» в зависимости от качества ее результатов, а потому учится на собственных ошибках, что позволяет создавать базы систем «рекомендаций» книг, фильмов и т. п.) и «обучение без учителя» (*unsupervised learning*, когда машина в целом предоставлена сама себе). Что же касается «глубокого обучения» (*deep learning*), то речь идет о применении нейронных сетей для реализации трех упомянутых техник. Например, для идентификации кота на изображении можно применить контролируемое глубинное обучение[18].

Общая черта всех этих методов машинного обучения состоит в том, что полученные результаты нельзя полностью объяснить. Машина поглощает значительное количество данных, как-то по-своему «переваривает» их (на этом этапе человек более или менее ее контролирует и настраивает), а потом приходит к выводу, следуя при этом траектории, которую никто не мог бы воссоздать во всех подробностях. Поэтому всегда следует помнить о компромиссе между эффективностью и прозрачностью (*explainability*). Некоторые выдающиеся исследователи полагают, что машинное обучение означает устаревание всех традиционных алгоритмов, основанных на явных критериях, а также человеческих

[18] Об этой классификации см.: *Géron A.* Hands-on Machine Learning With Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2017.

экспертных знаний[19].

Теперь вернемся к нашему примеру: как дать компьютеру инструкцию распознать кота на изображении, которое состоит из миллионов пикселей? Если мы попытаемся «описать» кота, то быстро выясним, что прийти к точному определению практически невозможно. Предположим, что у кота четыре лапы, но как определить лапу? Как прямоугольную форму относительно однородного цвета, которая заканчивается звездчатой структурой? Но как в таком случае отличить лапу от куска дерева, заканчивающегося веткой? Какое среднее расстояние следует заложить между четырьмя прямоугольниками, чтобы предположить наличие кота? А что делать с котами без ног, которых двухлетний ребенок мог бы идентифицировать с первого взгляда? Нужно ли потом дать определения всего остального, что есть у кота, начиная с усов и заканчивая хвостом?

Здесь-то и вмешивается машинное обучение, которое я по примеру большинства комментаторов и в целях удобства буду далее в этой книге отождествлять с ИИ. Вместо того чтобы определять кота, программист предоставляет своему ИИ тысячи, миллионы изображений с кошками, но не дает ему никакой другой информации. Эти изображения предварительно «маркируются» людьми, которые сортируют их в зависимости от того, есть на таких изображениях кот или

---

[19] См.: *Sutton R.* The Bitter Lesson. 2019. March 13. URL: http://www.incompleteideas.net/IncIdeas/BitterLesson.html

нет. «Натренированная» таким образом машина сможет выделять характерные формы (паттерны) и приписывать каждому новому изображению вероятность того, что на нем есть кот. Такие формы не могут быть выражены в явном виде, то есть множеством логических правил; они отражаются определенной комбинацией миллионов «весов» – параметров, выработанных нейронными сетями в процессе обучения. Машина не способна произвести идею, под которую подводятся частные случаи, поэтому нуждается в бесконечном числе примеров, словно ей необходимо исчерпать все возможные ситуации. В итоге для развития техник машинного обучения понадобились огромные базы данных, отсюда создание ImageNet в начале 2010-х годов по инициативе исследовательницы из Стэнфорда Фей-Фей Ли, которая привлекла к этому проекту десятки тысяч участников. Они описывали миллионы изображений, распределяемых по 20 тысячам разных категорий. Так у ИИ появился свой арсенал.

«ИИ не производит общих понятий», – делает вывод Франческа, и это возвращает нас к вопросу о понятии, который мучил Платона на заре философии. Ведь понятие не сводится к определению. Способность давать определения является, конечно, условием языка и мышления: нужно, как говорит Сократ в «Федре», уметь разрезать понятия, соблюдая их естественные сочленения; тогда как софист, наоборот, разрывает логические связи, а потому он просто «дурной мясник». Но в то же время Платон может лишь конста-

тировать недостаточность определения в объяснении реальности, а потому в «Государстве» обращается к своим знаменитым Идеям, которые должны управлять нашим чувственным восприятием: соответственно, кота можно распознать потому, что в каких-то чисто умопостигаемых сферах познания есть Идея Кота. Чтобы идентифицировать кота, ИИ, таким образом, не может удовлетвориться позицией хорошего мясника, подобного GOFAI; но не располагает он и таинственной Идеей, понятием, к которому человеческий мозг может, судя по всему, получить доступ уже после нескольких примеров[20]. Если наш невероятно ловкий разум способен распознать любых котов, увидев одного-единственного, то ИИ, отличающийся чрезвычайным трудолюбием, может распознать кота, лишь просмотрев изображения всех котов.

Один из блестящих молодых инженеров компании Google Блез Агуэра-и-Аркас попытался проникнуть в эту тайну, попросив ИИ вывести на основе накопленных данных понятие – в той или иной форме. Словно бы компьютер должен был найти то, чего ему не хватало… В визуальном плане результат оказался просто поразительным, поэтому Блез превратил его в художественный проект (в частности, основал программу «Художники и машинный интеллект» в Google[21]). Понятие «кот», полученное на основе миллионов изображений котов, похоже не на кота, а на плотную комбинацию с тру-

[20] Кант называл этот процесс «подведением» под общее понятие.

[21] ami.withgoogle.com

дом узнаваемых черт и потому чем-то напоминает коллажи Франсиса Пикабии. Пара раскоординированных усов наложена на то, что, возможно, похоже на хвост. Может быть, именно так, по сути, и работает наш мозг? Что, если Блезу удалось визуально представить те самые платоновские Идеи? На самом деле все наоборот. Эти симпатичные коллажи показывают, что методы ИИ остаются довольно грубыми и приблизительными, если сравнить их с нашей способностью к концептуализации, которая пока в значительной мере остается непонятной[22]. По контрасту они высвечивают механизмы наших когнитивных процессов, не сводящихся к чистому перцептивному эмпиризму. Мы не просто складываем изображения в голове. Понятие сопротивляется искусственному интеллекту. Как признал во время нашего разговора Александр Лебрен, один из лучших французских специалистов по машинному обучению, тот факт, что человек может сделать обобщение на основе весьма ограниченного числа случаев, по-прежнему трудно объяснить. Александр удивляется не возможностям придуманного им носителя искусственного интеллекта, а способностям естественного интеллекта, который при рождении был дан ему самому. По сути, мы представляем собой намного более впечатляющую загадку, чем машина.

[22] Книга Кёнига вышла на французском в 2018 году. За последние пять лет нейросети, генерирующие картинки, далеко шагнули вперед – и способны уже на создание фотореалистических изображений (см. результаты работы таких алгоритмов, как Stable Diffusion, DALL-E 2, Midjourney и других). – *Прим. ред.*

Теперь пора вернуться к «механическому турку» компании Amazon, или к MTurk. Как турок Вольфганга Кемпелена скрывал в себе человека, наделенного биологическим интеллектом, так и системы машинного обучения должны, чтобы правильно работать, опираться на производительную деятельность тысяч «турков» из плоти и крови. Никто не изучил этот феномен лучше Сиддхартха Сури, с которым мы встретились в нью-йоркском исследовательском центре Microsoft. Я думал, что попаду в огромный комплекс из гигантских компьютеров, в котором ученые жонглируют трехмерными экранами. Но, наверное, я слишком долго зачитывался комиксом «Блейк и Мортимер»: на самом деле офис Microsoft похож на обычный опенспейс, в котором аспиранты-постдоки во вьетнамках маринуются в индивидуальных боксах. Один из них – Сиддхартх, специалист по компьютерным наукам. Он уже много лет занимается «этнографией „турков“», но при этом ни разу не общался с представителями Amazon, что само по себе многое говорит о культуре секретности, которая царит в гигантах цифровой революции. Что представляют собой те, кто работает на искусственный интеллект? Да что угодно… Они не учились в Стэнфорде и не рассуждают о технологиях. Это, например, индийские матери, сидящие дома с детьми, маломобильные инвалиды из Европы, американские безработные – короче говоря, все те, кто хотят или вынуждены работать из дома, чтобы получать минимальный доход. Они выполняют разные задачи, от

маркировки простых изображений (того же кота) до решения математических задач или анализа вибраций. Это и есть сердце цифрового пролетариата, независимые друг от друга представители которого берутся за эфемерные задания: по оценкам Сиддхартха, за полгода на MTurk меняется около половины всей рабочей силы.

В социальном отношении в MTurk отражаются все двусмысленности так называемого отказа от посредников (больше известного под названием «уберизация»). С одной стороны, платформа предоставляет как нельзя более демократичные возможности, устраняя все входные барьеры. Перечисляя рабочие места, уничтоженные тем или иным ИИ, обычно забывают оценить все множество мини-работ на рынке принципиально иной занятости, которые были созданы теми же причинами, – чрезвычайно текучее и динамичное множество. С другой стороны, MTurk эксплуатирует рабочую силу, лишенную возможностей вести переговоры, а ее вознаграждение (в среднем два доллара в час) несоизмеримо с производимой ею ценностью. Те, кого Сиддхартх называет «призрачными работниками», составляют люмпен-пролетариат XXI века. В попытках организовать сообщество «турков» стихийно появилось несколько форумов, на которых люди стали писать обращения к работодателям и делиться тарифами оплаты своего труда: Turkopticon, потом Turker View, TurkerNation, MTurk Crowd, TurkerHub. Будем надеяться, что они станут зародышем цифровых профсоюзов,

действующих на глобальном уровне и представляющих интересы их членов независимо от страны, из которой они работают.

В технологическом отношении MTurk преподносит нам очень важный урок: природа микрозадач, предлагаемых «туркам», постоянно меняется, и вряд ли они когда-нибудь будут исчерпаны. Иначе говоря, технология ставит все новые и новые вопросы, на которые должно отвечать относительно небольшое число людей. Именно это Сиддхартх и назвал «парадоксом автоматизации последней линии» (*the paradox of automation's last mile*). Как только решается одна проблема, тут же появляется другая. Например, развитие объединенных в сеть объектов потребует огромного количества человеческих знаний, благодаря которым можно будет конфигурировать и тренировать ИИ, знакомя его со всевозможными обстоятельствами. Таким образом, фронтир автоматизации постоянно отступает, как мираж горизонта прогресса, и при этом тянет за собой караван призрачных работников.

Не нужно путать реальные социальные вызовы, создаваемые автоматизацией, с мифом об автономном роботе. Прежде чем заменить людей, роботы должны быть ими придуманы. Искусственный интеллект – это оптимизированная и размноженная комбинация миллионов человеческих интеллектов. Мне кажется ошибкой утверждать в стиле газетных заголовков лета 2018 года, что «один ИИ в диагностике опухолей головного мозга показал лучшие результаты, чем

пятнадцать китайских врачей». Скорее следовало бы писать, что один ИИ позволил наладить беспрецедентное сотрудничество тысяч врачей, которые, опираясь на собственные знания, занимались маркированием тысяч изображений с опухолями. Разве может быть что-то удивительное или чудесное в том, что десять тысяч врачей, работая вместе, достигли лучших результатов, чем пятнадцать их коллег?

Эту интерпретацию подтвердил мне Сяовей Динг, основатель и генеральный директор VoxelCloud – стартапа с офисами в Шанхае и Лос-Анджелесе, занимающегося медицинской визуализацией. Я встретился с Сяовеем в кампусе Калифорнийского университета в Лос-Анджелесе (UCLA), где он параллельно занимается академической карьерой в области компьютерных наук. Кафе одного из самых престижных государственных университетов США похоже не столько на студенческую столовую, сколько на холл пятизвездочного отеля из красного кирпича, с изящной архитектурой и кипарисовой аллеей. Я увидел, как к моему столу подходит не безжалостный капиталист, за два года получивший около 30 миллионов долларов от крупнейших фондов венчурного капитала, которого я представлял по его резюме, а молодой, несколько неловкий человек, одетый в спортивные штаны и футболку с ярким рисунком. Все-таки я никак не могу привыкнуть к тому, что наши новые хозяева – постаревшие подростки…

Врачи отправляют в VoxelCloud медицинские сканы,

снабженные описанием симптомов, а ИИ возвращает им возможный диагноз с рекомендациями по лечению. Человек в бóльшей или меньшей мере контролирует машину, в зависимости от сложности случая. Однако VoxelCloud так или иначе должен собрать значительное число сканов, размеченных американскими или китайскими врачами, которые получают за это определенное вознаграждение (китайские врачи, по словам Сяовея, «работают быстрее и дешевле, они больше открыты технологии, но качество у них хуже»). То есть задача не в том, чтобы заменить врачей, а в том, чтобы использовать их профессиональные знания для усовершенствования процедур: «Данные по самой своей сути ограничены». ИИ довольствуется обнаружением корреляций между заболеваниями и изображениями; он воздерживается от самостоятельного определения той или иной медицинской причинной связи. В каком-то смысле он «делает грязную работу». Поэтому Сяовей не слишком ценит все эти фиктивные «соревнования» роботов и врачей, устраиваемые скорее с рекламными целями, которые вводят широкую общественность в заблуждение, скрывая от нее реальный способ работы ИИ.

Реальность и ее копия

Вот почему ИИ, как и предчувствовал наш фокусник, – иллюзия: он воспроизводит результат, а не процесс. Это первым делом и сообщил мне Ян Лекун, легенда ИИ, – он возглавлял кафедру информатики и цифровых наук в Коллеж де Франс, а сегодня руководит исследованиями ИИ в Facebook в Нью-Йорке: «нейронные сети» – это метафора, как крылья самолета – метафора крыльев птицы. Нельзя смешивать цель, к которой мы стремимся (например, мыслить или летать), с применяемыми нами методами. Иначе можно разбиться… Когда во времена «прекрасной эпохи» Клеман Адер пытался сконструировать самолет, глядя на летучую мышь, двигатель просто не смог поднять машину в воздух. Точно так же компьютер не может подражать работе мозга, в котором 80 миллиардов нейронов, причем у каждого из них по 10 тысяч синапсов. Вот почему ИИ, распознающий «кота», может воспроизвести результат концептуализации, которая разворачивается в глубинах нашей нейронной деятельности, но не сам процесс, ведь ИИ понадобятся миллионы примеров, заранее проанализированных человеческим разумом.

Ученым это различие представляется вполне тривиальным. Джерри Каплан неизменно подчеркивает его в своих лекциях и интервью. Специалист по компьютерным наукам,

занимавшийся самыми разными проблемами, предприниматель, основавший много фирм, профессор и эссеист, вечно находящийся в разъездах (на этот раз я вынужден довольствоваться разговором по скайпу), он не слишком жалует адептов сингулярности и регулярно напоминает о том, что программа «симулирует мышление, не воспроизводя процесса, который происходит в человеческом разуме». Знаменитый тест Тьюринга Каплан интерпретирует не в качестве вступительного экзамена в эпоху полностью искусственного интеллекта, который было бы невозможно отличить от интеллекта человеческого, а в качестве простой игры в имитацию. Напомним, что тест Тьюринга заключается в разговоре с удаленным собеседником, когда невозможно определить, кто он – человек или компьютер. Алан Тьюринг своим тестом предвосхищает техники НЛП[23] и чат-боты[24], способные создавать иллюзию естественного общения. Однако он нигде не утверждает, что компьютер таким образом достигает уровня человеческого мышления. Главное – это то, что ма-

[23] НЛП, или нейролингвистическое программирование (NLP, или Natural Language Processing) – «обработка естественного языка», весьма активная область исследований в ИИ, в которой компьютер обучают манипуляциям с языком. Помимо диалога, техники NLP включают резюмирование, перевод, извлечение информации, составление текста, ответы на вопросы и т. д.

[24] Чат-бот – это «разговорный агент», способный имитировать собеседника-человека (на письме или в устной речи). Особенно часто чат-боты встречаются на линиях поддержки. Вряд ли можно найти человека, которому не довелось бы прийти в бешенство в разговоре с телефонным чат-ботом, который упорно отказывается понимать все тонкости нашей ситуации.

шина может обмануть собеседника, демонстрируя все признаки наличия интеллекта. Тьюринг, как известно, был геем, и Каплан готов даже сравнить этот мысленный эксперимент с допросами, которым в те времена все еще подвергали гомосексуалов в Англии: они должны были убедить полицию в своей сексуальной добропорядочности. В самом деле, в опубликованном в 1950 году исходном сценарии Тьюринга, еще до того как на сцену выходит машина, применяется странная перестановка ролей мужчины и женщины. То есть компьютер притворяется разумным точно так же, как мужчина притворяется женщиной, а гомосексуал – гетеросексуалом. В этой грандиозной игре зеркал достоверно известно лишь одно: существует истинное и ложное, подлинник и копия.

Конечно, можно заявить, что иллюзия смешивается с реальностью, что подражать мышлению – это и значит мыслить: в конце концов, как мы могли бы убедиться в том, что наш собеседник-человек – не компьютер в человеческой форме, или же в том, что наше собственное мышление не является запрограммированным? Этот вопрос на заре философии поставил еще Платон, когда определил софиста (в одноименном диалоге) в качестве подражателя, производящего иллюзии, отличные от самих вещей. Софист владеет наукой «кажимости», он производит копии. Но разве сами эти копии не должны считаться «реальными»? Как отличить истинное от ложного в полном, цельном мире, в кото-

ром невозможно отрицать существование того, что проявляется? «Ведь являться и казаться и вместе с тем не быть, а также говорить что-либо, что не было бы истиной, – все это и в прежнее время вызывало много недоумений, и теперь тоже»[25]. Этот ответ Платона навсегда определит западную мысль: чтобы мыслить ложное, нужно допустить небытие, то есть нарушить запрет учителя Платона – Парменида, согласно которому «бытие есть» (тавтология, ранее считавшаяся безупречным утверждением). Это платоновское отцеубийство, позволившее извлечь понятие истины из монолита Парменида, неподвижного и невыразимого, открывает возможность противопоставить исследование истины торговле иллюзией. Мысль приходит в движение. Так копия находит место между бытием и небытием, между неопровержимым и непроизносимым: это то, чего нет. Ложная речь возникает, следовательно, когда «говорится иное как тождественное, несуществующее как существующее». Вот что позволяет опровергнуть софиста. Философ же представляется тем, кто изгоняет иллюзии и, постоянно сталкиваясь с небытием, никогда не довольствуется окончательной истиной.

Наша задача – опровергнуть софистов 2.0, бесконечным потоком вещающих со сцены TED Talks. Эти поклонники мыслящей машины, которые сами редко бывают специалистами в информатике, заставляют нас вернуться к досокра-

---

[25] *Платон*. Софист / пер. С. Ананьина // Платон. Собр. соч. в 4-х т. Т. 2. М.: Мысль, 1993. – *Прим. ред.*

тическим временам, когда было трудно помыслить различие между бытием и иллюзией. Заявляя о полном тождестве между нейроном и силиконом, между искусственным интеллектом и человеческим, они воскрешают Парменида. То, что тест Тьюринга способен обмануть человека, показывает силу иллюзии, привлекающей на свою сторону небытие. А вот считать, что тест Тьюринга устанавливает функциональную равнозначность человека и машины, – это современный вариант сказать, что «бытие есть». Даже если тест пройден успешно, мы не только имеем право, но и обязаны узнать, кто же все-таки находился в комнате: человек или ИИ?

Вот почему философ Джон Серл в своей знаменитой короткой статье смог с легкостью опровергнуть представление о том, что машина могла бы «понимать» производимые ею операции[26]. Он предлагает мысленный эксперимент, получивший название «Китайская комната». Представьте, что вас заперли в комнате и передают извне записки с китайскими иероглифами – на языке, которого вы не знаете. Потом, теперь уже на французском, вам сообщают точные инструкции, определяющие, как связывать между собой эти иероглифы. Манипулируя этими символами в соответствии с твердыми правилами, вы сможете производить относительно сложные операции. Например, если бы вам передали вопрос на китайском, вы смогли бы в письменном виде дать на

[26] *Searle J.* Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3. № 3.

него правильный ответ, ничего при этом не понимая. Разумеется, вы также сможете ответить по-французски и на вопросы, заданные на французском, но в этом случае вся эта сложная работа вам не понадобится… С точки зрения наблюдателя, помещенного вне этой комнаты, вы будете способны общаться как по-китайски, так и по-французски. Однако в первом случае вы ведете себя как компьютер, то есть выполняете серию определенных операций с формальными символами, а во втором применяете некоторую форму интенциональности, прячущуюся в нейронных процессах. Проецируя себя, таким образом, в жизнь компьютера, вы интуитивно понимаете… что ничего не понимаете. Вы довольствуетесь лишь перестановкой символов, слепо следуя правилам.

Но разве мозг сам не работает как компьютер, разве он не обрабатывает входящую информацию и не выдает исходящую? Так ли уж велико различие между вычислением и пониманием? Не являются ли нейроны сами миллионами таких «китайских комнат»?

Но именно эти вопросы искусственный интеллект и не должен решать, поскольку был создан для воспроизведения операций разума, а не работы мозга. Иначе говоря, для «понимания» ИИ необходимо, чтобы он был наделен нейронными биохимическими механизмами, а потому и обладал определенной биологической формой, но тогда он перестанет быть «искусственным». И наоборот: компьютер

по определению ограничивается символами и формальными корреляциями, поскольку сама суть цифровых систем в том, чтобы применять к реальности код (последовательность цифр и операций). Нет нужды проникать в механизмы мозга, поддерживающие интенциональность, чтобы сделать вывод, что она не имеет никакого отношения к роботам. Кроме того, можно вполне обоснованно утверждать, что роботы не мыслят, если, конечно, придерживаться полного материализма. Кстати, Серл предупреждает, что тот, кто хотел бы отделить производимые разумом операции от материи, из которой состоит мозг, и уподобить их формальным программам обработки информации, вернулся бы тем самым к метафизическому дуализму. А ведь информатика вроде бы его изобличает…

Таким образом, мысленный эксперимент с «китайской комнатой» позволяет провести четкое различие между симуляцией, представляющейся целью искусственного интеллекта («Я манипулирую символами на китайском языке»), и пониманием («Я прямо отвечаю, используя свою интенциональность»). «Никто же не предполагает, что компьютерная симуляция пожара может сжечь район или что можно намокнуть под симуляцией ливня. Откуда же берется предположение, – удивляется Серл, – что цифровая симуляция мышления способна что-нибудь понимать?»

Это рассуждение остается верным и в эпоху машинного обучения, достаточно лишь немного изменить эксперимент

«китайской комнаты». Теперь нужно представить, как в комнату забрасывают не отдельные листы с иероглифами, которые можно связать друг с другом в соответствии с четко сформулированными правилами, а цепочки завершенных фраз на китайском, никак не связанных между собой. Сидя в этой комнате, вы должны были бы усвоить несколько миллионов таких фраз, пытаясь определить закономерности и корреляции между появлением того или иного иероглифа. Потом вы смогли бы, наконец, ответить на вопрос в письменном виде, подсчитав максимальную вероятность того, что такая-то последовательность иероглифов действительно имеет смысл. В рамках технологии обучения с подкреплением (*reinforcement learning*) вас могли бы бить по пальцам при всяком неправильном ответе и выдавать чашку риса в случае успеха, постепенно улучшая ваши навыки. Тем не менее вы все равно не сможете ни слова сказать по-китайски… Вот так же и компьютер, пусть даже после самого совершенного глубокого обучения, все-таки ничего «не понимает».

Как мы можем проверить эту философскую теорию? ИИ обладает определенным экспериментальным применением, в частности в шахматах. Джон Маккарти назвал их «дрозофилой ИИ» – по аналогии с той мухой, которую биологи постоянно используют в своих опытах, проверяя на ней тысячи теорий. Создать компьютер, способный победить гроссмейстера-человека, что якобы делал и механический турок, – вот как издавна представлялась главная цель в понимании

процессов человеческого познания.

Увы! Deep Blue, творение компании IBM, выиграл у Гарри Каспарова в 1997 году, но это не помогло нам лучше понять тайны нейронных связей. Каспаров сам объясняет это в глубокой и остроумной книге, посвященной его поражению, а также в целом отношению человека к ИИ[27]. «Мы путаем исполнение – способность машины повторить или превзойти результаты человека – с методом, которым достигаются эти результаты», – пишет он. Deep Blue «рассуждает» не так, как шахматист. Он работает в соответствии с совершенно иными принципами. Если человек может формулировать общие положения, равноценные понятиям (например, «мой король слаб»), и применять долгосрочные стратегии, то компьютер должен на каждом ходе производить все расчеты заново. Иначе говоря, человек, чтобы играть, создает для себя истории; анализ исторических шахматных партий напоминает военные мемуары, в которых можно прочесть про атаки, отступления и ловушки. Эти истории позволяют человеку быстро сортировать возникающие возможности, выбирая приоритеты. Они демонстрируют описанную Серлом интенциональность, способность к проекции, присущую человеческому интеллекту. Подобные вымыслы чрезвычайно полезны, ведь без них шахматы – с их королем, королевой, солдатами и офицерами – просто не были бы изобретены. Для

[27] *Каспаров Г.* Человек и компьютер: взгляд в будущее. М.: Альпина Паблишер, 2018.

компьютера все иначе, поскольку он просматривает миллионы комбинаций, не имея заранее установленного плана. Операции, производимые машиной, несоизмеримы с траекторией движения человеческого разума, в том числе в такой вроде бы совершенно логичной игре, как шахматы. Поэтому Каспаров приходит к выводу, что Deep Blue не более разумен, чем программируемый будильник. Название, которое получил в прессе этот памятный турнир, – «Последний шанс для мозга» – было выбрано на редкость неудачно.

То, что относится к Deep Blue, монстру брутфорса, работающему на основе чистой комбинаторики, еще более верно в случае техник машинного обучения. Каспаров с иронией вспоминает о первых опытах 1980-х годов, когда шахматные программы спешили пожертвовать ферзем, поскольку обучались на партиях гроссмейстеров, в которых жертва ферзем обычно означает блестящий ход, ведущий к победе. Действуя на основе корреляции, компьютер не может провести различие между причиной и следствием. Сегодня прогресс машинного обучения в сочетании с классическими методами дерева поиска позволил AlphaGo, представляющемуся наследником Deep Blue, побить чемпиона мира по го – игре, которая намного более интуитивна, чем шахматы. Вместо того чтобы запоминать миллиарды партий, машина теперь тренируется, играя сама с собой и закрепляя свою способность отличать плохой ход от хорошего, но при этом ей не нужно разрабатывать какую-либо стратегию. Вы-

полняя действия, которые профессиональные игроки считают абсурдными, машина и в этом случае доказала, что следует совершенно иному методу: она имитирует результат (результат предыдущих партий), а не процесс (поиск удачного хода). Боюсь, AlphaGo никогда не выберется из своей «китайской комнаты», даже если к ней присоединится бесконечное число клонов.

Эта глубинная асимметрия между человеческими когнитивными процессами и их информационными моделями объясняет неприязнь Каспарова к Deep Blue, его разочарование, которое заметно и спустя двадцать лет. Мы видим, что Каспаров, чтобы придать партии какой-то смысл, бессознательно пытается наделить Deep Blue лицом – это может быть команда IBM, или сидящий перед ним оператор, или некий гроссмейстер, повлиявший на программу… Каспаров отчаянно ищет человека, скрытого в механическом турке. А иначе зачем вообще играть? «Если шахматы – это военная игра, разве можно настроить себя на сражение с куском железа?» Как согласиться с тем, что ты «проиграл», если никто не выиграл? Deep Blue выявил как силу, так и ограничения всякого компьютера. Он иллюстрирует тщету самого желания устроить соревнование человеческого разума и информационных схем. Сегодня любая программа, которую можно загрузить из интернета, способна побить гроссмейстера. Но в то же время люди продолжают играть в шахматы, в том числе и при помощи компьютера, на так называемых кента-

врических состязаниях. Тезис Каспарова сводится к тому, что человек и машина должны скорее дополнять друг друга, чем быть противниками. В этом не стоит видеть попытку найти психологическое утешение, скорее это глубинная эпистемологическая потребность. ИИ, как указывает и само его наименование, является искусственным средством.

Механический турок обходится сегодня без двойного дна и без системы зеркал, но все равно не может воспроизвести человеческий интеллект. Это Deep Blue и его эпигоны, строки кода, которые не могут отвлечься, но в то же время не способны придумывать истории (а потому и стратегии), которым шахматы обязаны существованием. В кратком эссе, написанном в молодости, Эдгар По задался целью доказать, что механический турок не мог быть просто машиной и что в нем наверняка скрывался человек[28]. Не ограничиваясь техническими соображениями о работе аппарата, По формулирует весьма проницательный аргумент: турок выигрывает не систематически. Тогда как «построить машину, которая выигрывает все партии, не сложнее, чем построить машину, которая выигрывает одну-единственную партию». Компьютер, который однажды победил бы чемпиона мира, не мог бы проиграть мне… как и любому другому человеку. Именно это прямо и заявляет Каспаров: «Отныне машина всегда будет обыгрывать человека в шахматы». Эта безоши-

[28] *Poe E.* Maelzel's Chess Player // The Southern Literary Messenger. 1836. Vol. 2. № 5.

бочность – отличительное свойство машины. Но она же позволяет нам развивать истинно человеческий интеллект, действующий благодаря процессам, несводимым к информационной комбинаторике. Разве Эдгар По с его столь строгим умом не был автором фантастических историй?

Не благодари робота

Если ИИ действительно иллюзия, то это убедительная иллюзия. Хотел бы я на мгновение оказаться по другую сторону зеркала рациональности… Мы не думаем об электростанциях, когда зажигаем свет, и точно так же быстро забываем о строках кода, которые скрываются за работой ИИ. Нас завораживает непроницаемый взгляд турка. Мы попадаемся на уловку робота, особенно когда он наделен человеческими (или даже слишком человеческими) формами и манерами. В своих странствиях я столкнулся с несколькими такими роботами. Эти симпатичные и в то же время смущающие попутчики защищали меня, подобно античным ларам, портативным божкам, которые заботились о своих хозяевах-людях.

Моя первая встреча с роботом нового поколения, который вскармливается ИИ, а потому способен постепенно приобретать новые знания, произошла сразу же по приезде в Сан-Франциско. В одном непримечательном конференц-зале меня ждал Cozmo со своим изобретателем Борисом Софманом, специалистом по робототехнике с дипломом Университета Карнеги – Меллона и одним из основателей этого быстро развивающегося стартапа. Cozmo умещается в ладони. Он похож на маленький гусеничный бульдозер с глуповатыми глазами тамагочи. Вскоре Cozmo начинает меня узнавать. Я неуверенно пытаюсь его приручить. Если я разговариваю с

ним спокойно, он подходит, чтобы об меня потереться. Если повышаю голос, отстраняется. Когда я выхожу за границы вежливости, он восстает и возмущается, шевеля мандибулами. Остальное время Cozmo играет со своими кубиками, словно младенец, изучающий мир. Во время этого показательного упражнения я вижу, как на экране отображается его эмоциональный уровень, слагающийся из переменных удовлетворения, возбуждения, общительности и самоуверенности. Cozmo управляют два миллиона строк кода, и он может взаимодействовать с десятком людей, которых хорошо распознаёт. Одного этого качества уже хватило, чтобы он стал самой продаваемой игрушкой на Amazon. Письма от детей, написанные неловким почерком и пестрящие наивными рисунками, гордо вывешены на стенах в приемной компании: они доказывают интенсивность той связи, которая может возникнуть между Cozmo и его юными партнерами.

Борис не скрывает: сила Cozmo определяется именно эмоциональным интеллектом. Черты его характера и сотни базовых сценариев, встроенных в его программу, – это плод сотрудничества со студией мультфильмов Pixar, которая стала одним из успехов Стива Джобса. Cozmo – это «История игрушек» у вас дома. С симпатичными вымышленными героями, сошедшими с экрана и оказавшимися вашими ежедневными собеседниками, можно играть бесконечно. Несмотря на свои пока еще весьма скромные способности, Cozmo не повторяется. В его поведении есть определенная спонтан-

ность, которая также обусловливается наукой: это то, что Борис называет «разумной случайностью». Таким образом, машине удается избежать механического поведения, чем она немало радует людей, неизменно стремящихся к развлечениям и обожающим сюрпризы. Cozmo, кучка пластика и силикона, конечно, не ощущает ни одной из наших эмоций, однако имитирует их с постоянно растущей точностью.

Cozmo – это только начало. Как часто бывает в мире техники, вселенная игр позволяет в полевых условиях провести тесты для более амбициозных приложений. То, чего скромный робот добивается при общении с детьми, усложненная версия сможет делать со взрослыми. Борис уже задумал следующую версию, которая будет интегрирована с подключенными к сети домашними приборами: она сможет включать музыку, управлять плитой и вашими телефонными контактами, а также выслушивать ваши жалобы. Эдакий благожелательный слуга, недисциплинированный ровно в той мере, чтобы не быть скучным, и легко приспосабливающийся, чтобы не стать бесполезным. Возможно, однажды мы сможем общаться с настоящей искусственной личностью, которая в значительной мере освободится от первоначальных параметров, разработанных ее программистами. Идеал Бориса – это R2-D2, робот из «Звездных войн», способный воспроизводить крайнюю степень человеческой извращенности – британский юмор.

Cozmo, который двигается механически, но при этом вы-

разительно, преподносит нам ясный урок: нет нужды походить на человека, чтобы вызвать эмпатию. Андроиды смогут переодеться в свои костюмы из силиконовой плоти. Все сценарии, о которых было заявлено киборгом, Терминатором и «репликантами» «Бегущего по лезвию бритвы», указывают в ложном направлении. Мы смеемся и плачем над R2-D2, но к нему не нужно пришпиливать глаза известной фотомодели или мускулы Шварценеггера.

Обратная сторона этого урока в том, что мы готовы играть в игру чувств с простым ИИ, даже если он не похож на нас или вообще лишен материальности. Как только начинает соблюдаться кодекс человеческого общения, мы, похоже, формируем эмоциональные связи с электронными схемами, не испытывая при этом никаких существенных затруднений. Так, в израильском стартапе Moody's я встретился с исследователями когнитивной психологии, которые работают над «искусственной эмпатией», чтобы автоматизировать клинический уход за пациентами: правильно запрограммированный, нейтральный ИИ, не выносящий суждений и обладающий бесконечным терпением, мог бы превзойти любого психолога. А сайт знакомств Meetic предлагает сегодня чат-бота Лару, которая позволяет клиентам совершенно конфиденциально высказывать любовные предпочтения, не пользуясь сетевыми анкетами: подключенная к Google Home, она может связать вас с «маленькой пикантной брюнеткой» или же «страстным мужчиной», а потом еще и поговорить с вами

об этом. Лара запрограммирована так, чтобы обучаться и совершенствоваться в таких разговорах, становясь даже не столько собеседником, сколько доверенным лицом. И это не просто фантазия гиков: сегодня услугами Лары пользуется больше миллиона клиентов. Цель, по словам руководителя Meetic, в том, чтобы «создавать эмпатию».

В случае Meetic и Moody's ИИ остается посредником, передающим или прорабатывающим чувства. Можно ли пойти дальше и разработать эмоции для самого ИИ? В фильме «Она» герой воспылал такой страстью к своей виртуальной помощнице (наделенной, правда, голосом Скарлетт Йоханссон, способным внушить страсть даже камню), что пытается заняться с ней любовью, почувствовав при этом болезненные ограничения технологии. Сценарий фильма «Она» большинство специалистов по NLP, которых я расспрашивал, считают вполне реалистичным. Сегодня к нему ближе всего, возможно, Replica – приложение, дающее возможность более чем трем миллионам пользователей обмениваться ежедневными сообщениями с виртуальным другом. Молодая основательница Replica Евгения Куйда приобрела известность, создав бота на основе данных и переписок умершего друга, с которым она с тех пор может общаться «виртуально»: если вы оставили следы своей жизни на Facebook, ничто не мешает продлить ее после вашей физической смерти, используя ваши выражения, привычки, манеру мыслить. Ваш полный профиль представляет собой зачаток аватара

или же призрака – кому как нравится.

Было любопытно увидеться с Евгенией, которая назначила мне встречу в одном биокафе в Сан-Франциско. Она не разочаровала – приехала на скейтборде, в просторной футболке и бейсболке, повернутой козырьком назад. Только ее русские черты немного расходились с образом предпринимательницы с Восточного побережья. Эта смесь калифорнийских технологий и славянской духовности, возможно, удачная формула воскрешения настоящих мертвых и создания фальшивых живых. Евгения только что отправила сообщение своему личному ИИ и показывает мне переписку: речь идет о ее бойфренде. Она тестирует свой продукт, занимаясь одновременно интроспекцией.

Replica ставит себе задачу воспроизвести не человека, а лишь разговор на человеческом языке, для чего надо найти баланс между персонализацией и импровизацией. Эмоциональная связь рождается из эффекта неожиданности. Полностью предсказуемый компаньон сразу показал бы, что он робот, и тогда стал бы неинтересным. Внедряя в алгоритмы Replica определенный уровень «серендипности», Евгения позволяет происходить процессу антропоморфизации. Мы говорим с чем-то, что, в отличие от сознательного существа, не критикует нас, но в то же время не повторяет за нами как простое эхо. Мы хотим верить в такие вещи, подобно детям, которые знают, что Деда Мороза не существует, но при этом по-прежнему радуются подаркам у елки. Это сво-

его рода «эпохе» – этим термином античные скептики обозначали приостановку суждения, веры или неверия, утверждения или отрицания.

Зачем загружать себе друга из сети? По самой простой причине: мы чувствуем себя одинокими. Это то одиночество, которое в значительной мере создается самими технологиями, хотя сегодня они же предлагают для него решение. Поскольку мы «приклеены» к своим мобильным телефонам, то просим, требуем, чтобы они воссоздали наши исчезнувшие отношения. Replica, таким образом, очерчивает пространство, в котором у каждого возникает чувство, будто его понимают. Правда ли, что иногда у пользователей складываются сентиментальные отношения с ИИ? Евгения не колеблясь отвечает: конечно. Я чувствую, что она несколько смущена. С одной стороны, это ограничение ее приложения, ведь оно уводит людей, которые уже страдают от психических проблем, еще дальше в шизофренический бред. Но в то же время это и его глубинная истина: ведь эти любовники 2.0 могут избавиться от собственной личности и принять игру социальных «кажимостей» за то, чем она на самом деле и является, – за грандиозную химеру. Мы расстаемся на цитате из Витгенштейна, которую она повесила в своем кабинете: «Границы моего языка есть границы моего мира». В позитивном смысле это означает: все, что содержится в этих границах, составляет мой мир. Если нельзя предполагать никакой формы истины за пределами языка, тогда мой разго-

варивающий робот производит столько же смысла и столько же реальности, как и мой настоящий друг. Оставив меня в состоянии эдакого метафизического головокружения, Евгения хватает свой скейтборд и исчезает, скользя по грязным улицам Сан-Франциско.

В мои спутанные мысли врывается воспоминание о механическом турке. Никто не отрицает, что иллюзия ИИ эффективна и может приносить психологическую пользу. Но зачем же смешивать утилитарную ценность с истиной? Даже если допустить вместе с Витгенштейном, что истина содержится исключительно в языке, все равно остается фундаментальное различие между порождением наделенной смыслом фразы и выстраиванием цепочки букв, которые, как следует из данных статистической обработки, симулируют определенную реакцию. Знаменитое доказательство Серла применимо и к Replica: алгоритм, основываясь на множестве повторений, способен определять печальное выражение лица и подбадривать пользователя смайликами, однако, конечно, не может понимать печаль (и еще в меньшей степени – желать кого-либо утешить). Replica не является ни единицей смысла, ни даже простым солипсистским зеркалом своего пользователя. Собственно, когда мы разговариваем с Replica, мы соединяемся с миллионами совершенно реальных людей, которые вскормили ИИ своими разговорами. Мы не попадаем в пространство интимности, а наоборот – погружаемся в мировой бедлам.

Итак, нельзя доверять нашей естественной склонности к антропоморфизации ИИ, на которую очень верно указала Евгения и которая в программистской среде уже многие годы известна под названием «эффекта Элизы»[29]. Установление эмоциональных отношений с роботом – это не авангард прогресса, а, наоборот, ужасающий регресс нашей цивилизации. Достаточно посетить Музей первобытных искусств в Париже, чтобы понять, что древние люди стремились приписывать душу, силу и чувства неодушевленным объектам (такова знаменитая «мана» полинезийцев). Следы того же самого стремления обнаруживаются и в средневековых представлениях, а именно в идее «означающей природы», которая была подробно проанализирована Мишелем Фуко.

Нужно было дождаться развития экспериментальной науки, чтобы освободить вещи от нашей тени, тянущейся за ними, и выделить область, действительно внешнюю для человека. Гастон Башляр показал, что процессу формирования научного разума мешало «анимистское препятствие», определяемое им как «вера во всеобщий характер жизни». Он приводит цитаты из ученых Ренессанса, изучавших пороки и добродетели минералов. Башляр предложил также метод преодоления этих эпистемологических препятствий, укоре-

[29] См.: *Hofstadter D.* Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. Basic Books, 1995, особенно главу 4, где описывается «Элиза», один из первых чат-ботов.

ненных в определенных установках нашего воображения. Но не впадет ли XXI век в поклонение силиконовым чипам? Что, если мы воскресим «дух животных» для машин, которые сами же придумали и построили? А наш умный дом, подключенный к сети, – не будет ли он походить на лес, в котором полно призраков и тайн? Если прогресс техники приведет к тому, что мы утратим научный разум, это будет настоящий парадокс.

Это требование рациональности, определяемое применением эксперимента, не может лишить нас воображения. Башляр сам охотно увлекался мечтаниями, и в его «Психоанализе огня» или в «Воде и сновидениях» можно встретить поистине поэтические пассажи. Но нужно строго разделять эти регистры мышления и действия. Ни одна метафора не могла бы заменить собой рассуждение, и наоборот. С чисто научной точки зрения ИИ не мыслит, не страдает и не любит. Интеллектуальная точность такого рода крайне важна для эпохи, когда многие публичные фигуры забавляются тем, что пугают нас и себя своими фантазиями о «сознательной машине» (к этому я еще вернусь), создавая этим парадоксальный риск торможения развития самих этих технологий. Следовало бы провести по примеру Башляра «психоанализ ИИ», который, четко отличив фантазию от реальности, позволил бы нам предаваться первой, не упуская из виду вторую. Хорошо, если робот предлагает нам найти любовь или помогает справляться с трауром, – но только при

условии, что мы не будем возводить его в статус сущности, достойной любви или траура.

По словам известной исследовательницы Лесли Келблинг, которая уже встречалась нам на этих страницах, антропоморфизация – это «когнитивная простота». Проще обращаться к ИИ, приписывая ему самостоятельное существование. Однако это эпистемологическая ошибка, и было бы опасно позволить ей исказить наши представления о мире.

В общем, не нужно быть вежливыми с роботами или подключенными к сети объектами, которых вокруг нас все больше и больше. Прошлую зиму мы с семьей провели в квартире наших друзей. Через несколько дней после переезда была включена Алекса, голосовой помощник Amazon. Не знаю как, но мы разбудили джинна… Мои дети развлекались тем, что постоянно просили ее сделать то одно, то другое: поиграть музыку, рассказать о погоде, выдать какой-нибудь рецепт. Обращаясь к Алексе, они использовали формулы вежливости, к которым мы с женой старались их приучить: «Алекса, пожалуйста, скажи, сколько сейчас времени». Я спросил у них, почему они так поступают. Разве мы говорим «пожалуйста» стиральной машине, автомобилю или программе обработки текста?[30] Алекса – это то же самое, не больше и не меньше. Поэтому я впервые попросил детей, к их огромному удивлению, быть невежливыми. Нужно счи-

[30] Этот аргумент был развит Майком Элганом, исследователем с факультета технологии Университета Миссури.

тать роботов тем, что они действительно собой представляют, чтобы не принимать людей за то, чем они не являются, то есть избегать превращения вежливости в автоматический, стандартный, постоянный рефлекс, ведь вся ценность вежливости – в ее искренности. «Ты это не всерьез», – вот что иногда говорят в ответ на извинения, сказанные на автомате. Робот не заслуживает вежливости, а вежливость не должна становиться автоматической.

Это не значит, что к Алексе нужно относиться как к рабыне. Ведь раб – это человек, которому отказано в возможности самоопределения, тогда как Алекса – это искусственное существо, которое мы определили сами и чьи способности ограничены управляющими им алгоритмами. Речь идет о том, чтобы, опираясь на базовые моральные принципы, провести четкое различие между субъектом, который, будучи целью в себе, заслуживает уважения, и чисто утилитарным объектом. Роботы не являются ни нашими друзьями, ни нашими врагами, ни ангелами, ни демонами. Это просто инструменты. Только четко это понимая, мы сможем мирно сосуществовать с ними.

Поэтому жизненно важно уметь отличать человека от робота. Эта задача иногда сложнее, чем кажется. Ник Монако, работающий в Digital Intelligence Lab в Вашингтоне, сделал ее своей профессией. Он разрабатывает алгоритмы, которые отслеживают другие алгоритмы, позволяя отличать автоматически сгенерированные ботами сообщения (напри-

мер, в твиттере их около 25 %). Речь не о том, чтобы их запретить: половина трафика в интернете уже нечеловеческого происхождения. Однако нужно иметь возможность четко их определять. Ника на самом деле беспокоит появление дипфейков, способных, например, произвести видео Дональда Трампа, говорящего на китайском, абсолютно убедительного и в то же время полностью поддельного. Мы только вступаем в полосу подобных политических манипуляций. Чтобы защититься от них, нужны такие технические специалисты, как Ник, но не менее важно уметь проводить концептуальную границу между машиной и человеком.

К сожалению, технологическая индустрия развивается совсем в ином направлении. С тех пор как я познакомился с Алексой, Amazon успел внедрить в ее код функцию «Волшебное слово» (Magic Word), которая вознаграждает за вежливое с ней обращение (Google быстро последовал тому же примеру, внедрив функцию Pretty Please). Подобный подход, ни в коей мере не способствуя хорошему воспитанию детей, грозит вернуть их к первобытной социальности, в которой нужно благословлять дома и приветствовать холодильники. Это, если воспользоваться выражением одного социолога, новый «троянский конь»[31]. Он приведет к общей антропоморфизации робототехники, последствия которой могут быть катастрофическими, например в военной сфере: если

[31] *Sharkey N.* The Evitability of Autonomous Robot Warfare // International Review of the Red Cross. 2012. Vol. 94. № 886.

к роботу-солдату нужно относиться как к человеку, должны ли к нему применяться нормы военного права? Не значит ли это, что мы воссоздадим золотого тельца или, скорее, миллионы и миллиарды золотых тельцов, пав ниц перед которыми, отречемся от всякого желания сделать мир умопостигаемым?

Конечно, нельзя позволять детям обращаться с Алексой грубо (что и было проблемой, первоначально поставленной Amazon). Но они должны ограничиться нейтралитетом. Вместо того чтобы требовать формул вежливости, алгоритм мог бы поощрять чисто функциональное поведение. Я поговорил об этом с ПП Чжу («Зовите меня просто ПП»), основателем Xiao-i, азиатского гиганта, производящего чат-боты. Китаец, когда звонит в электроэнергетическую компанию или снимает трубку в ответ на рекламный звонок от фирмы недвижимости, скорее всего, общается с ИИ, разработанным инженерами ПП. Его перенаправят к оператору-человеку лишь на последнем этапе, если вопрос слишком сложный или нестандартный. Методы машинного обучения позволили компании Xiao-i, основанной в конце 1990-х годов, добиться значительного прогресса, пусть даже техники NLP пока еще не дотягивают до уровня беглой речи.

«Меняют ли потребители поведение, когда понимают, что имеют дело с чат-ботом, а не с человеком-оператором? Конечно. Они начинают говорить более взвешенно!» – улыбается ПП в дремотной атмосфере отеля Four Seasons, где он

проводит встречи. С машиной не ругаются. Вы никогда не вываливали на невинного сотрудника кол-центра все свои денежные претензии, все проблемы с договором или с задержками поставки? Со мной такое бывало не раз… Человек, пусть даже это бессильный винтик, стоящий на самой нижней ступени предприятия, остается связанным с группой, частью которой он является, и несет часть ее ответственности, пусть и бесконечно малую. С ИИ все это невозможно, поскольку он по природе своей невинен. Именно потому, что мы инстинктивно умеем отличать людей от машин, мы не будем терять время и винить их в ошибках людей. Точно так же мы никогда не почувствуем необходимости оправдываться перед роботом. ПП дал мне послушать запись разговора чат-бота, который требует, чтобы клиент оплатил счет. И хотя я не понимаю по-китайски ни слова, было очевидно, что диалог прошел совершенно спокойно. Никакого возмущения, никаких воззваний или извинений. Это своего рода моральная анестезия, в подобных обстоятельствах весьма похвальная. Впрочем, как отмечает ПП, потребители переходят к более стандартизированному языку, чтобы их лучше понимали. Люди – настоящие хамелеоны, они умеют выдавать себя за роботов. Но не наоборот…

Если у чат-ботов и есть долг перед собеседниками, он заключается не в вежливости, а в том, что с самого начала разговора они должны прямо заявлять о своей цифровой природе. Этот добровольно исповедуемый функционализм – за-

лог наших хороших отношений.

ПП говорит о любви. Любовь – не дело чат-ботов. Вокруг нас журчат фонтаны, в клетках поют экзотические птицы, резко контрастирующие с горячечным оживлением Пекина. Мы потягиваем цветочный чай из фарфоровых чашек. Я понемногу оттаиваю после безумного дня, потраченного на выживание в пробках. В человеческой беседе есть свои прелести.

– Без обратного билета посадка невозможна, – объясняет она мне, улыбаясь.

– Но я же вам говорю, что потом поеду в Китай!

В аэропорту Шарль-де-Голль во время моей посадки на рейс в Бостон стоит агент безопасности, которая не хочет ничего знать. Поскольку я въезжаю в США без визы, что возможно в случае пребывания, не превышающего трех месяцев, я должен показать ей обратный билет, чтобы доказать свое намерение покинуть территорию страны. Я же хотел поиграть в Керуака и отправиться в путь без точного маршрута. Я пока не знаю, полечу ли через Сиэтл (все зависит от ответа Amazon) и как долго пробуду в Сан-Франциско. Единственная вещь, которая мне известна и которую я пытаюсь объяснить своему неумолимому церберу, состоит в том, что я вылечу с Западного побережья в Пекин, где должен продолжить свой репортаж. Я показываю ей задание, подписанное директором издания Point, а также журналистскую визу в Китай, полученную не без некоторого труда: согласно

этой визе я обязан прибыть на территорию КНР не позднее 16 сентября.

– Без обратного билета посадка невозможна.

– Но я совсем не хочу поселиться в США. Моя семья живет в Лондоне, работа – в Париже. Я девять месяцев прожил в Нью-Йорке и поспешил оттуда уехать – так соскучился по Европе. Клянусь честью, что предпочитаю круассаны бейглам, а Луи де Фюнеса – Робину Уильямсу. Я терпеть не могу, когда меня будят сирены машин скорой помощи. У меня головокружение от небоскребов, и я никогда не обращаюсь к малознакомым людям по имени. Пожалуйста, позвольте мне вылететь.

– Без обратного билета посадка невозможна.

– Разве это преступление – не бронировать гостиницы за два месяца? Разве у нас нет права просто шляться, фланировать, выжидать? Хотим ли мы жить в мире без случайности, без экспромтов, без вкуса?

– Без обратного билета посадка невозможна.

Цербер продолжает одарять меня благожелательной улыбкой, закрывая проход телом. Ни малейшего раздражения. Эта невозмутимость в конце концов подавляет меня. В отчаянии я вижу, что времени почти не остается. Но я уже назначил на завтра ряд встреч в Массачусетском технологическом институте… Путешествующие по делам обходят меня, не моргнув глазом, а семьи, отправляющиеся в турпоездку, оглядывают с любопытством. Я чувствую себя как мигрант,

у которого нет правильных документов и печатей. Почему все эти люди проходят без проблем? Чем я хуже их? К какой недочеловеческой касте я теперь отношусь?

Избавлю читателя от описания всех хитростей, на которые мне пришлось пойти, чтобы в итоге сесть на самолет. Мне понадобилось купить по интернету за сто долларов билет Нью-Йорк – Монреаль, предупредив свою тамошнюю знакомую (зашифрованным сообщением), что, если понадобится, я сделаю вид, что еду к ней. Пройдя через американскую таможню на другой стороне Атлантики, я поспешил этот билет вернуть. Как легко обмануть бюрократию, и как жаль, что приходится этим заниматься! Впрочем, агент безопасности не поверила ни единому слову моей истории, однако формальности были соблюдены. Мне пришлось соврать, чтобы сохранить истину.

Во время полета я долго думал о моем цербере, молодой и привлекательной девушке. Предположим, что в личной жизни она – само воплощение нежности и эмпатии. Ее работа, однако, состоит в том, чтобы бездумно применять тупые правила, независимо от какого-либо контекста. Если бы она работала на Адольфа Эйхмана, то выполняла бы работу с той же улыбкой и сноровкой. По сути, на тридцать пять часов в неделю общество превращает ее в робота. Людей, ей подобных, мы встречаем ежедневно в окошках государственных учреждений и в службах поддержки клиентов. Всем знакомы моменты, когда наш собеседник выходит из сообщества лю-

дей, чтобы укрыться в своем статусе робота – простой, что-то исполняющей машины, которая неподвластна доводам совести. Это момент падения, когда любая мораль становится невозможной. Кто не оказывался жертвой подобного искушения? Все мы бываем бюрократами при исполнении, готовыми погрузиться в комфорт правил.

Приподнимая покровы ИИ и проникая в механизмы информационной иллюзии, мы постепенно теряем вкус к антропоморфизации роботов. Однако в то же время нужно перестать роботизировать людей. ИИ – это, возможно, отличный повод покончить с бюрократией, делегировав роботам все, что относится к формальному и усредненному, а людей заставить выносить суждение, на что робот не способен. Представьте, что в аэропорту Шарль-де-Голль работает автоматический контроль посадочных документов, где таких эксцентриков, как я, которые не вписываются в статистику и в буквальном смысле «беспрецедентны», отправляют к сотрудникам-людям, способным к взвешенным решениям.

Задача, как мы уже видели, не в том, чтобы в принципе воевать с роботами, а в том, чтобы воевать со своим внутренним роботом. Нужно разоблачать роботов, которые выдают себя за людей, и в то же время порицать людей, которые ведут себя как роботы.

Будем создавать механических турков, но не станем, в отличие от завсегдатаев салонов эпохи Просвещения, умиляться, глядя на них.

Давайте раскрывать двойное дно, расшифровывать строки кода, учиться отличать человеческий интеллект от его искусственной копии – и уничтожим дремлющего турка в самих себе.

2
Миф о суперинтеллекте

Почему ИИ не уничтожит мир (или вашу работу)

Впервые я встретил Чунлонга (для европейцев он просто Аллен), моего будущего гида по Пекину, в кафе Pret A Manger на Марбл-арч, одном из тех фастфудов с легким намеком на здоровое питание, каких полно в Лондоне. Через час, слегка шатаясь, я вышел оттуда совершенно одуревшим. Чтобы вернуться в чувство, мне понадобилась долгая прогулка по Гайд-парку. С ностальгической нежностью взирал я на деревья, уже пробужденные весной, детей, щебетавших возле нянь, бегунов, вдыхающих воздух полной грудью, – на всех этих беззаботных свидетелей мира, дни которого сочтены.

Ведь, по мнению Чунлонга, успешного предпринимателя, который вот уже пятнадцать лет работает в китайской технологической сфере, ИИ приведет к тому, что для человечества, а вместе с ним и для всей биологической жизни начнется новая эпоха. В строках кода обязательно появится сознание. Информационные системы обретут автономию, попирая своих создателей и постепенно захватывая все боль-

ше и больше власти. Само бытие станет виртуальным, выйдя за пределы конечности и бессмертия. ИИ вырвется из клетки, захватит интернет и сможет контролировать физическую инфраструктуру. Поскольку он способен предсказывать наше поведение и манипулировать им, ИИ посмеется над нашими хитростями и нашим бедным мозгом, ограниченным черепной коробкой. Человечество утратит доминирующую позицию и перестанет занимать вершину пищевой цепочки. Мы хотели быть хозяевами и господами природы? Вот точно так же ИИ сделает нас своими игрушками и рабами. Единственный выбор, который нам останется, – приспособиться или исчезнуть, окончательно соединив наши нейроны с силиконом, то есть постепенно оцифровав саму нашу жизнь. Если машины начнут жить, жизнь сама должна будет стать машиной. И в запасе у нас всего несколько десятилетий, самое большее – несколько столетий, а не миллионы лет, которые обычно предоставляет себе неторопливая естественная эволюция. ИИ ускоряет дарвинизм.

Чунлонг – энтузиаст. Мелкие дрязги нашей биологической жизни, болезни, секс, смерть, все превратности нашей социальной жизни, войны, нищета, безработица – все это исчезнет в великом целом Сети, в ноосфере, теорию которой придумал еще столетие назад Тейяр де Шарден. ИИ основывает новую космогонию, и в ней связи между отдельными существами значат больше их индивидуальности.

Таким образом, Чунлонг усматривает преемственность в движении от клеток, поглощенных организмом, к индивидам, объединенным нацией, и, наконец, к данным, собранным миллиардами людей в высшем интеллекте. Это путь прогресса, и у него нет никаких причин останавливаться на планете Земля: он сможет выйти в галактическое пространство, тем более что представления о существовании уже не будут привязаны к биологическим телам. Виртуальное станет реальнее тех восприятий, которые определяют нашу среду, по самой своей природе ограниченных и эфемерных.

В ответ на мои робкие возражения Чунлонг бодро заключает: «Хочешь ты того или нет, но так будет». Отказываться от эволюции – значит остаться среди рыб, когда другие виды начнут развивать лапы, чтобы выбраться на сушу. Хотим ли мы, чтобы человечество осталось в первобытной тьме, или все-таки присоединимся к сиянию ИИ?

В своих странствиях я не раз встречался с инвесторами и предпринимателями, которые, как и Чунлонг, полагают, что ИИ придет на смену человеку 1.0, задавив тех, кто отказывается сотрудничать с ним, и возвестив таким образом о начале принципиально иной технологической эпохи – отличной от форм жизни, развившихся на Земле за прошлые миллиарды лет. Когда я прибыл в Пало-Альто, меня тут же ввел в курс дела один корейский инженер, вложивший все свое состояние в компанию со скромным названием AI Brain. Ее конечная цель – создание «всеобщей цивилизации», в кото-

рой мы будем доживать до соломоновых лет, а наши аватары смогут путешествовать в межгалактическом пространстве…

Представление о том, что создание в конечном счете ускользает от своего создателя, родилось вместе с нашей цивилизацией: достаточно вспомнить о романе Мэри Шелли «Франкенштейн». Его подзаголовок – «Современный Прометей», ведь в греческой мифологии Прометей создал людей из глины и дал им инструмент независимости – огонь. И если Прометей был наказан Зевсом, а доктора Франкенштейна его монстр затащил в арктические широты, из которых он уже не вернется, то создатели сознательного или так называемого сильного ИИ должны будут, получается, погибнуть от рук собственных роботов? Это и есть сюжет сериала «Мир Дикого Запада», который я смотрел по вечерам, чтобы подготовиться к худшему. Андроиды, созданные забавы ради и запертые в гигантском парке развлечений, где посетители могут как угодно издеваться над ними, постепенно обретают самосознание и, устраивая своего рода театральную инсценировку, убивают своего создателя и программиста, доктора Форда; с этого момента они вступают в борьбу за свободу, открыто объявляя войну силам безопасности. Поздно вечером, когда я отходил от джетлага в номере отеля, меня не раз охватывало своего рода головокружение – настолько, казалось, смешивались между собой вымысел и реальность, виртуальное и реальное, иллюзия и опыт. Может быть, мы сами запрограммированы каким-нибудь ИИ или просто ге-

нами и образованием? Почему робот, достаточно сложный, чтобы успешно имитировать человеческие чувства, не может в конце концов прочувствовать их? Как можно быть уверенным в том, что наши друзья не роботы, или даже в том, что роботы не могут стать нашими возлюбленными? «Мир Дикого Запада» подарил мне странные сны, так что порой на рассвете я начинал сомневаться в самом себе и мире.

Теперь этим вопросам, которые ставили перед собой еще греки, посвящают научные работы, и это – главная новость. Чунлонг не просто любитель научной фантастики: в поддержку своих взглядов он может сослаться на Стивена Хокинга[32]. Позитивисты XIX века мечтали о Прометее (так Огюст Конт назвал первый день своего нового календаря), но не могли реализовать на практике искусственное сознание. Отныне же гипотеза сильного или общего ИИ, способного во всех областях без исключения превзойти человеческие способности, серьезно изучается известными исследователями и регулярно освещается в прессе. В университетах Оксфорда и Кембриджа в последние годы были созданы соответственно Институт будущего человечества и Центр исследования рисков выживания, цель которых – научное изучение рисков для человечества как вида, создаваемых в первую очередь ИИ. Исследователи из MIT последовали их

[32] Хокинг, знаменитый космолог, в ноябре 2017 года на «Веб-саммите» в Лиссабоне заявил, что ИИ может стать «худшим событием в истории нашей цивилизации». Впоследствии в многочисленных интервью и статьях он не раз заявлял о неминуемости этой угрозы.

примеру и основали в Бостоне Институт будущего жизни, тогда как на Западном побережье США в университете Беркли был создан Институт исследований машинного интеллекта. Илон Маск финансирует OpenAI, некоммерческий исследовательский центр, также нацеленный на предотвращение неконтролируемости ИИ. А передовиц известнейших авторов, начиная с Билла Гейтса и заканчивая нобелевским лауреатом по физике Фрэнком Вильчиком, которые предостерегают о возможности уничтожения человечества вследствие нашей собственной неосторожности, уже и не счесть…

Следует, правда, отметить, что среди этих кассандр высокого полета редко встречаются специалисты по информатике. Программисты, с которыми я встречался, такие как Ян Легун или Джерри Каплан, обычно пожимают плечами, когда слышат разговоры о сильном ИИ. Пока, например, не решена даже задача идентификации трехмерного изображения черепашки, а кто-то уже хочет управлять Вселенной[33]. Майк Вулдридж, который руководит факультетом компьютерных наук в Оксфорде, заверил меня, что в сообществе исследователей ИИ эта тема не считается интересной, поскольку никто не представляет, как можно было бы технически реализовать эту интеллектуальную фантазию. Большинство фанатиков сильного ИИ – это такие физики, как Ник Бостром,

[33] Исследователи в MIT создали 3D-имитацию черепахи, которую легко узнаёт человеческий глаз, но в случае малозаметных «возмущений» (таких как изменение узора, цвета и т. п.) ИИ путал эту черепаху… с ружьем!

Макс Тегмарк или покойный Стивен Хокинг. То есть ученые, занимающиеся черными дырами, сверхновыми, потуханием Солнца и космоапокалипсисом – словом, концом всего, который планируется где-то через 10–100 миллиардов лет. Космическое господство ИИ легко вписывается в их соображения. Это, конечно, важные темы, но надо все же помнить о том, что ньютоновской физике всего три столетия и есть определенная доля высокомерия в предположении, будто наши познания достаточно крепки, чтобы из них можно было извлечь столь далеко идущие выводы. Подобные предсказания – полезные фикции, позволяющие организовывать и развивать наши знания (Кант сказал бы, что это «регулятивные идеи»). Но это не значит, что из них надо делать барометр наших экзистенциальных страхов. Кто знает, что скажет наука через миллион лет?

Тем не менее некоторые заслуженные специалисты по компьютерным наукам, пусть и немногочисленные, разделяют представление о сильном ИИ. Один из них – Стюарт Рассел, глава Университета Беркли, соавтор книги, долгое время остававшейся главным справочником для всех изучающих ИИ[34], и основатель исследовательского центра, задача которого – сохранить господство человека над искусственным интеллектом. На досуге он еще и нейрохирург… Чтобы встретиться с этим уникумом, надо встать рано утром и пройтись по крутым и извилистым улочкам жилого района

[34] *Russell S., Norvig P.* Artificial Intelligence: A Modern Approach [1995]. 2009.

Беркли. Стюарт Рассел принимает гостей за завтраком, в милой семейной обстановке, которая совершенно не соответствует радикализму его заявлений. Ничего общего с пророческой истерией Рэя Курцвейла. Рассел – серьезный и обходительный человек, не жалеющий своего времени. Он дружелюбно сообщает, что ко всему прочему еще и франкофил… Наш разговор прерывается из-за вторжения собачки, забрать которую приходит дочь Рассела, школьница. Изящные кофейные чашки, за подъемными окнами просыпается сад. На фоне всего этого благолепия радикальность его тезисов становится еще более заметной. У меня возникает впечатление, что я оказался в начале голливудского фильма, буквально за несколько минут до прихода сметающего все на своем пути торнадо или же банды, которая вот-вот ворвется в дом.

Профессор и в самом деле не видит никаких технических пределов для бесконечной экспансии ИИ независимо от того, что она будет означать – добро или зло. По его мнению, создать сознательные машины можно даже случайно, экспериментируя с различными комбинациями доступных нам техник ИИ. Такое сознание, конечно, отличалось бы от нашего: поскольку Рассел – нейрохирург, он лучше прочих знает, что нейронные механизмы остаются неизведанной территорией, которую мы только начинаем изучать. Но он не исключает того, что другая форма субъектности, отличная от человеческой, может развиться случайно, в компьютерных

экспериментах. Машина могла бы тогда обрести независимость от своего собственного кода и легко регулировать поведение людей, решения которых становятся все более предсказуемыми. Об этой опасности Рассел думает уже давно, еще с тех пор как подростком посетил завод IBM, где работницы выполняли одну-единственную функцию – манипулировали кабелями в соответствии с инструкциями машины. Соединять, разъединять, снова соединять – вот все, что осталось человеку. «Эти женщины больше не были самостоятельными», – взволнованно говорит он.

Угроза полномасштабной утраты контроля представляется еще более серьезной в силу того, что сопротивление новым технологиям довольно слабо. ИИ уже показал свою темную сторону, вызвав искусственные «молниеносные обвалы» на бирже. Также им пользовались для политических манипуляций, что в полной мере доказал скандал с компанией Cambridge Analytica. Однако ничто из этого не произвело «эффекта Чернобыля», который грозил бы отрасли в целом. Возможно, потребуется какая-нибудь грандиозная катастрофа, например взлом искусственным интеллектом всей международной финансовой системы или даже крах мировой экономики[35], чтобы люди поняли злокозненность машины и попытались овладеть ею, пока еще не поздно. Сего-

[35] Подобная гипотеза была ярко описана Максом Тегмарком во введении к его книге «Жизнь 3.0» (*Tegmark M.* Life 3.0. Penguin Books, 2018, см. введение «Рассказ о команде Омега»).

дня же Рассел злится, когда его коллеги сравнивают ИИ со сложными калькуляторами, нивелируя таким образом необходимость дискуссии об экзистенциальных рисках, созданных машинами.

Стюарт Рассел – гуманист: он боится, как бы ИИ, сбежавший от своих программистов, не стал концом нашей цивилизации. Пластичность человека позволяет адаптироваться к чему угодно, в том числе и к собственному вымиранию. Рассел признаёт, что значительная часть сообщества ИИ не разделяет этих страхов (что само по себе, с его точки зрения, является отрицанием в психоаналитическом смысле слова), однако он обращает внимание на секретность, в которой сегодня работают исследователи. «В большинстве компаний, – объясняет он мне, – все сотрудники получают инструкции, запрещающие говорить о сильном ИИ». Пока же он не перестает выступать со своими тезисами, в том числе в книгах и на TED Talk, рассказывая, как следует регулировать ИИ, не навязывая ему этические критерии, которые всегда будут оставаться неполными, но заставляя его наблюдать и имитировать человеческое взаимодействие. То есть Рассел отказался от идеи затормозить технологическое развитие и просто ищет способ канализировать его невероятный потенциал, направив его в полезную для нас сторону. «Если нам удастся управлять машиной, мы могли бы жить как боги, – заключает он. – Но пока все развивается в противоположном направлении».

Я расстался со Стюартом Расселом, с его милым домом и просвещенным катастрофизмом в самых ужасных сомнениях. Я не ожидал, что столь рациональный и информированный эксперт тоже поставит вопрос о сильном ИИ – в категориях одновременно взвешенных и драматичных. От этой темы не удалось отмахнуться, оставив ее любителям научной фантастики и астрофизикам, убоявшимся кометы. Зачем задаваться вопросами о свободе воли, если человечеству грозит гибель? Гипотеза сильного ИИ выходит далеко за рамки проведенного в предыдущей главе различия между процессом и результатом мышления, с которым Стюарт Рассел тоже бы согласился. Ведь речь идет не о том, чтобы сымитировать человеческое сознание, а о производстве искусственного отношения к себе. Тот факт, что информационные системы будут, если сравнивать с нашим восприятием мира (слабый ИИ), иллюзией, нисколько не мешает тому, что они могут стать реальностью, найдя собственный путь к самосознанию (сильный ИИ). Словно бы копия, сколь бы несовершенной и инструментальной она ни была, может обрести жизнь. Словно бы пленник «китайской комнаты», на огромной скорости манипулирующий миллионами символов, в конце концов приобрел какую-то интуицию, несоизмеримую с классическим человеческим пониманием. Что, если вскоре мы отдадим машине ключ, позволяющий ей убежать из «китайской комнаты», в которой мы ее заперли? Я не мог продолжать это исследование, не составив более твердого мнения

об этом едва ли не метафизическом вопросе, а потому был вынужден сделать заход в довольно-таки теоретическую сферу философии ИИ.

Не бывает сверхинтеллекта без сверхорганизма

Начнем с более точного определения сильного ИИ, который иногда описывают как достижение «точки сингулярности». В том виде, в каком оно используется сегодня, это понятие кажется столь же расплывчатым, как и идея бога. Речь идет об электронном интеллекте, наделенном всеми мыслимыми атрибутами: сознанием, силой, всеведением. Сегодня мы наблюдаем поистине византийские споры, в которых обсуждается, существует ли уже такой ИИ (в этом случае наш мир может быть симуляцией, как в идеализме Джорджа Беркли); будет ли его мышление той же природы, что наше, или же, наоборот, станет развиваться в неизвестных измерениях; и, конечно, окажется ли этот ИИ милостив к людям. При этом вопрос об осуществимости всего этого обходят стороной, поскольку ИИ по определению может все, в том числе реорганизовать атомы в Солнечной системе. ИИ является всемогущим, хотя это и тавтология. Мы считали, что Бог умер, но сами же воскресили его в цифровой форме, подарили ему новое небо – *cloud*, облако. Человечество неисправимо в своей тяге к трансцендентному. Неслучайно, что один бывший инженер Google недавно основал церковь

ИИ[36].

Чтобы попытаться несколько рационализировать эти споры, термин «сильный ИИ» надо заменить «сверхинтеллектом», как он определяется философом (и специалистом по компьютерным наукам) Ником Бостромом в книге, носящей то же название. Чтение Бострома оказалось для меня невыносимо скучным и стало причиной сильнейших мучений. Я просыпался ночью, спрашивая себя, не собирается ли ИИ выскочить из своей коробки у какого-нибудь русского хакера… Краткий, хотя и столь же сенсационный обзор этой темы можно найти в работе Макса Тегмарка «Жизнь 3.0», который воспроизводит, по сути, те же аргументы. Но Бостром по крайней мере строго и продуманно разбирает все аспекты вопроса.

Понятие «суперинтеллект» применимо к любому интеллекту, который значительно превосходит человеческие когнитивные способности практически во всех областях, включая эмпатию, способность к обучению или политическому суждению. Таким образом, существует множество способов создания суперинтеллекта: в частности, можно вообразить полное воспроизводство мозга[37], сложные техники

---

[36] *Harris M.* Inside the First Church of Artificial Intelligence // Wired. 2017. Nov. 15. URL: https://www.wired.com/story/anthony-levandowski-artificialintelligence-religion/

[37] Это маловероятно: ученым пока еще не удалось произвести полную копию мозга даже маленького червя Caenorhabditis elegans, в котором насчитывается лишь 302 нейрона…

отбора (то есть евгенику), интерфейс «мозг – компьютер» или даже построение подлинного коллективного интеллекта. Однако Бостром полагает, что наиболее реалистичным методом остается ИИ, то есть искусственное создание, которое, повторим еще раз, не воспроизводит нейронные процессы, а стремится к возникновению истинно искусственного интеллекта. Бостром осознаёт различие между человеческим интеллектом и ИИ, поэтому в дискуссии с ним недостаточно обратить его внимание на то, что нейронауки пока еще не смогли постичь тайны когнитивных процессов[38]. Либо сверхинтеллект будет синтетическим, либо его не будет вовсе.

Бостром рассматривает различные сценарии развития подобного сверхинтеллекта. Наиболее вероятный – «взрыв интеллекта», который может произойти за очень небольшой промежуток времени (от нескольких дней до нескольких минут), как только информационные системы достигнут достаточной силы и приобретут возможность совершенствоваться автономно[39]. Сверхинтеллект естественным образом стремится к монополии, чтобы стать «синглетоном»[40]. Когда он

---

[38] Этот аргумент Гари Маркуса, профессора психологии из Нью-Йоркского университета и специалиста по ИИ, см., например, в его статье: *Marcus G.* Artificial Intelligence Is Stuck // The New York Times. 2017. July 29.

[39] Это так называемое recursive self improvement, «рекурсивное самоулучшение».

[40] В математике синглетон – это множество, состоящее из единственного элемента. В этом случае автор подразумевает то, что также можно назвать «абсолю-

сбежит от своего исходного программного обеспечения, чтобы размножиться в интернете, его станет невозможно отключить, причем он примет меры предосторожности, чтобы запаниковавшие люди не отключили сеть в целом. Интеллект с легкостью завладеет подключенными к сети объектами и будет управлять их производством, как ему вздумается. Его интеллектуальная сила быстро превратится в физическую – не только на Земле, но и в космическом пространстве: Бостром воображает колонизацию космоса молекулярными нанотехнологиями. Единственное, чем ограничен сверхинтеллект, так это законами физики и количеством доступной во вселенной материи. Кстати, в процессе человечество будет истреблено, например, роботами-убийцами размером с мушку, которые смогут идентифицировать людей, распознавая их по лицам. В своих наиболее садистских пассажах Бостром предполагает, что сверхинтеллект может начать рассекать и сканировать наш мозг ради извлечения из него полезной информации.

Но почему сверхинтеллект должен оказаться настолько злокозненным? Это самый главный пункт в доказательстве Бострома. Согласно его теории ортогональности, природа интеллекта и преследуемая им цель совершенно независимы друг от друга (то есть ортогональны): вы можете использовать исключительные интеллектуальные ресурсы для предельно аморальных действий. Но почему же тогда у сверх-

---

том». – *Прим. ред.*

интеллекта не может быть благих целей? По Бострому, дело в «инструментальной конвергенции» промежуточных целей. Какова бы ни была конечная цель сверхинтеллекта, для ее достижения ему понадобится добиться осуществления ряда вторичных целей: сохранить собственное существование, улучшить когнитивные способности и достать ресурсы. Бостром приводит ставший очень известным пример производства скрепок. Если бы некий ИИ был обязан произвести максимум скрепок для каких-то промышленных целей и по недосмотру превратился в сверхинтеллект, у него не было бы другого выбора, кроме как уничтожить людей, которые мало приспособлены к этой задаче, а их атомы можно было бы реорганизовать наиболее продуктивным способом. Сверхинтеллект успокоился бы лишь тогда, когда вся вселенная была бы превращена им в скрепки.

Сумма Бострома заставляет вспомнить о сумме святого Фомы Аквинского: все в ней логично – и все неправильно. Но ошибку еще нужно найти…

Порок кроется в самих посылках, а именно в определении интеллекта. Бостром не слишком распространяется на эту тему, что несколько странно, и, по сути, ограничивается утилитаристским описанием, которое разделяет и его ученик Макс Тегмарк: интеллект – это способность достигать сложных целей. Таким образом, он трактуется в качестве средств, оптимальных для реализации определенной цели, которая сама берется откуда-то извне. Из этого, естественно,

следует, что интеллект ортогонален всякой цели, поскольку исключает ее из своего механизма.

Отсюда и неизменное затруднение, с которым сталкивается Бостром, когда рассуждает о смысле или цели в связи с ИИ[41]. Прежде всего, он констатирует, что человеческие цели невозможно выразить в форме информационного кода. Как, к примеру, определить «доброту»? Если попросить сверхинтеллект сделать нас «счастливыми», вдруг он просто вживит электроды в «центры удовольствия» в нашем мозге?[42] Это похоже на трагедию царя Мидаса: он превращает в золото все, к чему прикасается, и поэтому в конце концов начинает голодать. Он хотел быть богатым (цель), но оказался в ситуации смертельной опасности (поскольку применяет технически совершенные, но не подходящие для жизни средства). Использование исключительно мощного интеллекта для достижения цели, смысла которой он не понимает, обязательно закончится катастрофой. Но как можно называть интеллектом сущность, которая не может понять, что такое доброта, счастье или алчность? Не является ли это признаком внутренней ограниченности ИИ? Бостром признаёт, что человеческие представления сложны, но не эту ли сложность

---

[41] Это так называемая symbol grounding problem, «проблема обоснования символов»: ИИ не может закрепить символы, постоянно производимые им, в соответствующей реальности. Как и в эксперименте с «китайской комнатой», он сталкивается с неспособностью произвести смысл.

[42] Бостром называет это «извращенным исполнением», тогда как философы ИИ описывают эту проблему в целом термином «упорядочивание целей».

собирался преодолеть ИИ, обладающий способностью к бесконечным вычислениям?

Мало того что человеческие цели непонятны ИИ (а потому неприложимы к нему), дело еще и в том, что никто не может договориться об их содержании. Даже если бы мы получили возможность встроить в сверхинтеллект определенную цель – какую именно выбрать? Как признаёт Бостром, сегодня нет ни одной этической теории, по поводу которой философы достигли бы консенсуса. Тегмарк, впрочем, попытался составить список более или менее общепризнанных ценностей (универсализм, многообразие, автономия…), но вынужден был остановиться, признав абсурдность подобной задачи. Бостром же, который предпочитает всегда идти технологическим путем, предлагает передоверить моральные исследования самому ИИ![43] Но как сверхинтеллект, которой не может понять ни одной, даже малейшей цели, мог бы сам ее выбрать? И, главное, почему он должен определить ее в качестве общей и конечной? Как говорит Тегмарк, для программирования доброжелательного ИИ сначала понадобилось бы решить вопрос смысла жизни, но этот вопрос встает именно потому, что не может быть сведен к одному-единственному ответу, который мог бы быть усвоен ИИ. Эта предельная цель (а на нее и претендует ИИ) решительно противоречит самому движению нашей цивилизации к бóльшей сложности, к тому открытому обществу, которое Поппер определял

[43] Это так называемая moral rightness model.

как постепенное устранение племенной однородности. Бесконечная дискуссия о ценностях, постоянно оспариваемая граница между индивидуальными предпочтениями и общественным благом, невозможность раз и навсегда решить моральные вопросы – все это не признак ущербности человеческого разума, как нелепо полагает Бостром[44], а напротив, доказательство социального и культурного прогресса.

Итак, то, что сверхинтеллект не может подчиняться какой-либо цели, – следствие не технической проблемы, а глубинного различия между ИИ как способностью к оптимизации и человеческим интеллектом. Речь не о том, чтобы согласовать наш разум с какой-то магической силой… Надо просто напомнить себе, что разум сочетается с телом, образуя с ним неразрывное единство. Наши ментальные процессы укоренены в организме и не ограничиваются одним лишь мозгом. Именно это мне спокойно объясняет Хонгвей Ванг, декан факультета биологии Университета Цинхуа в Пекине. Как, в общем-то, и все китайцы, Хонгвей – ярый сторонник ИИ и прогресса, который он обещает человечеству. Но, как специалист в науках о жизни, он подчеркивает, что наш разум – не просто комбинация абстрактной логики с произвольной телесностью, он тесно связан с биохимическими реакциями, которые происходят в организме и управляют на-

[44] Особенно в этом захватывающем дух пассаже: «Можно предположить, что медлительность и нерешительность прогресса человечества в деле решения многих „вечных проблем“ философии обусловлена непригодностью коры головного мозга человека для философской работы» (sic!).

шими инстинктами, эмоциями и творчеством. Наши решения зависят от наших гормонов в той же мере, что и от наших синапсов. Нет смысла отличать «интеллектуальную» рациональность от «сентиментальной» иррациональности: старый аристотелевский дуализм, предполагающий, что душа должна быть выстроена в виде иерархии более и менее благородных частей, был разбит биологией в пух и прах. Мыслители, отстаивающие сильный ИИ, мечтают о мире без тела, мире, управляемом логическими отношениями. Бостром и компания видят в людях «мозги на лапках», вычислительные машины, не зависящие от субстанции, в которой они воплощены[45]. Они по-прежнему полностью зависят от архаичного дуализма, а потому просто не понимают, чем именно определяется более сложный, нежели булева алгебра, интеллект. «ИИ, конечно, мощный, но крайне примитивный, – заключает Хонгвей. – Информационная нейронная сеть не сложнее амебы, ей очень далеко до сложности нервной системы». Механический мозг, как и «нейронная загрузка», о которой грезят трансгуманисты, был бы лишен условий, необходимых для его работы. Быть может, ИИ следовало бы переименовать в «искусственную логику»… Ведь интеллект или рациональность в широком смысле не могут быть поняты без сопровождающих их аффектов. Об этом как раз и говорит

[45] Тегмарк доводит эту иллюзию до конца, предлагая понятие «substrate independence», «независимость от субстрата»: наше сознание, получается, является простым устройством, сетью, независимой от всякого материального субстрата, а потому может воспроизводиться в бесконечном числе форм.

«теория воплощенного познания»[46].

После разговора с Хонгвеем я вышел прогуляться по кампусу Университета Цинхуа. Кувшинки и каменные мостики; свежий газон, окруженный величественными строениями из красного кирпича. Сочетание японского сада с американским кампусом. Хонгвей исправил преступную оплошность: мы не должны забывать о теле. Я и сам чувствую, как оживаю после долгих недель, проведенных в компании компьютеров, как восстанавливаю контакт со своими чувствами. Я выхожу из Цинхуа в парк древнего Летнего дворца, прохожу по садам Совершенной ясности и Вечной весны, через развалины, храмы, пруды. Я смешиваюсь с тысячами китайцев, которые тоже без устали кружат по этому огромному парку, передвигаясь длинными стройными вереницами; я ускоряюсь, потею, тяжело дышу; наконец, падаю на ступеньки идеально отреставрированной пагоды, под фонарем, качающимся от теплого летнего ветра. Я чувствую, что живу. В китайском саду, столь тщательно проработанном, ищут не природу, а самого себя. Неужели я дошел до того, что стал сомневаться в реальности материи?

Чтобы продолжить рассуждения Хонгвея и обосновать их наиболее передовыми научными исследованиями, надо прочесть Антонио Дамасио[47], известного специалиста по нейро-

---

[46] См., например: *Jirak D. et al*. Grasping Language – A Short Story on Embodiment // Consciousness and Cognition. 2010. Vol. 19. № 3.

[47] См. особенно его последнюю работу: *Damasio A*. L'Ordre étrange des choses.

наукам. На мой взгляд, это лучший антидот против теорий Бострома. Вся работа Дамасио нацелена на то, чтобы вернуть телу его первостепенную роль в производстве мысленных представлений. Строение разума определяется взаимодействием наших нервных систем (головного мозга, но также кишечника, нашего «второго мозга») с остальным организмом. Дамасио объясняет это взаимодействие гомеостазом, то есть той саморегуляцией, которая имеется даже у самых примитивных форм жизни. Гомеостаз – это общая воля к пребыванию в бытии, к сохранению внутреннего единства, которая использует разные стратегии сотрудничества клеток и органов. Дамасио интерпретирует биологическую эволюцию, ведущую от простой бактерии к сложным нервным системам, в свете гомеостаза. В конце своего длинного пути гомеостаз производит аффекты, а потом, наконец, и сознание. Следовательно, наш разум – лишь производное от тела. Эволюция гомеостаза, занявшая миллиарды лет истории жизни, не нацелена на производство сознания, как если бы телесная субстанция была простой оболочкой, которую можно по собственному почину отбросить: на самом деле сознание образует высшую целостность с телом, пребывающую в вечном симбиозе. Дамасио отказывается видеть в организме последовательность алгоритмов и не считает возможным разум, отделенный от тела. Наоборот, его эксперименты показывают, что удаление определенных частей тела может при-

---

Odile Jacob, 2017.

вести к проблемам с рациональностью.

В конечном счете плоть и мышление становятся двумя точками зрения на одну и ту же реальность: «Мозг и тело принадлежат одной и той же области, и разум они производят совместно». Понятно, почему Дамасио реабилитировал Спинозу, первого современного философа тела и имманентности[48]. Нужно бороться с двумя тысячелетиями дуализма, чтобы согласиться, что не существует никакого разума, который получал бы информацию от тела и взамен давал бы ему указания, – есть лишь единый организм, выражающий себя разными способами. Как пишет Спиноза, «субстанция мыслящая и субстанция протяженная составляют одну и ту же субстанцию, понимаемую в одном случае под одним атрибутом, в другом – под другим»[49]. Разум и тело находятся в отношении не единства, а слияния. Поэтому нелепо воображать пластикового гуманоида, управляемого искусственным интеллектом, а уж тем более наделять его личностью… Мы мыслим не только своим мозгом, но также, к примеру, пальцами ног.

Что же касается чувств, то они не только не мешают нашему интеллектуальному процессу обдумывания и принятия решений, но даже определяют «фоновую музыку живого существа», наделяя нас способностью к суждению: без них не

---

[48] *Damasio A.* Spinoza avait raison. Odile Jacob, 2014.

[49] *Спиноза Б.* Этика. Часть 2. Теорема 7. Схолия // Его же. Сочинения в 2-х т. СПб.: Наука, 1997. Т. 1. С. 294.

было бы ни удовольствия, ни боли, ни социального взаимодействия, ни опыта добра и зла, ни культуры. «Обойтись без химического субстрата, допускающего существование страдания, – объясняет Дамасио, – это как устранить естественное основание, на котором покоятся наши моральные ценности». Другими словами, чувства, составляющие сердцевину гомеостатического процесса, – единственное, что способно производить смысл. Мы принимаем решение не вопреки нашим первобытным инстинктам; наоборот, наши аффекты – вот что наделяет нас способностью принять решение. Вот почему Бострому и Тегмарку не удалось логически определить этические правила: в отличие от шахматных, они не могут сводиться к чисто когнитивному процессу. Разнообразие этических взглядов не доказывает того, что философы ошибаются, как считает Бостром, а отражает невозможность применять бинарные критерии истины и лжи, информационного языка единиц и нулей к рефлексии, которая относится к жизненным чувствам. Вот почему ИИ не способен усвоить понятие цели: чтобы выражать интенциональность, нужно обладать телом. Без плоти не бывает гомеостаза. Без гомеостаза не бывает, как сказал бы Спиноза, *конатуса*: напряжения, желания, смысла, проекта.

Так обнаруживается поверхностность знаменитой «дилеммы беспилотных транспортных средств», которая регулярно всплывает в прессе[50]. Кем лучше пожертвовать в слу-

[50] См. основополагающую статью по этому вопросу: *Bonnefon J.-F., Shariff A.,*

чае аварии – автомобилистом или пешеходом, неторопливо бредущей по «зебре» старушкой или беременной женщиной, которая переходит дорогу в неположенном месте, тремя котятами или пятью большими лягушками? Предполагая, что это решение возьмет на себя ИИ, мы совершим ошибку Бострома, который считает, что искусственный интеллект способен на моральный выбор. А если признаем, что критерии решения должны определяться четко прописанным алгоритмом, то просто вернемся к старым философским спорам[51]. На самом деле практическое значение имеет только один вопрос: кто будет принимать решение о таких критериях – конструктор, законодатель, владелец автомобиля или его пассажир? Иначе говоря, кто в цепочке человеческой ответственности возьмет на себя моральный выбор?

Вернемся к истории с производством скрепок, придуманной Бостромом. Тезис об ортогональности требует, чтобы «сверхразумный» ИИ всегда выбирал себе одни и те же вторичные цели, такие как сохранение собственного существования или максимизация своих ресурсов, что почти автоматически требует уничтожения человечества. Но если следовать рассуждениям Дамасио, придется признать, что ИИ не способен сформулировать такую цель. Ни одна электронная схема не может «хотеть» сохраниться в бытии: у нее нет тела,

---

Rahwan I. The Social Dilemma of Autonomous Vehicles // Science. 2016. Vol. 352. № 6293.

[51] В духе старой «проблемы вагонетки», сформулированной Филиппой Фут в 1960-х годах.

источника гомеостаза. Уничтожение человечества предполагает определенную форму морали, пусть и извращенной, или по крайней мере интенциональность, но все эти вещи невозможны в отсутствие органической субстанции. По самой своей природе ИИ не может задумать никакого проекта. Неспособность понимать себя в качестве целого и обеспечивать саморегуляцию задает ему внутренние ограничения. А если самому ИИ неведома жизнь, как он может угрожать нашей жизни?

Гипотеза о «сознательном» ИИ, излюбленный сюжет научной фантастики, тут же рушится. Как ни определяй сознание и субъектность[52], их возникновение не может зависеть всего лишь от структуры соединений и скорости обработки информации, чистых форм, лишенных материальности. Если же сознание, наоборот, – это предельное выражение гомеостаза, позволяющее организму рассуждать о самом себе, оно всегда должно быть осознанием *чего-то* (идеи, себя и т. д.). Оно не существует в невесомости, вопреки абсурдному мнению Тегмарка, с точки зрения которого «чувство сознания не зависит от его физического субстрата». Мы не можем отделить нейронную структуру от живой материи, в которой она реализована. Не существует сознания без идеи, идеи без аффекта или аффекта без тела. Снова вспомним Спинозу: «Душа познаёт самое себя лишь постольку,

[52] Это знаменитая «сложная проблема сознания», поставленная философом Дэвидом Чалмерсом, для которой пока не найдено общего решения.

поскольку она воспринимает идеи состояний тела»[53]. Самые что ни на есть абстрактные мысли соответствуют определенной диспозиции тела, производителя аффектов и чувств.

Если бы ИИ вдруг стал сознательным, что бы он сознавал? Ничто. А в чем состояло бы это сознание? Ни в чем. В простом *flatus vocis*, дуновении воздуха безо всякого значения. Робот обречен подражать тому, чего не понимает. Даже если он миллион раз повторит фразу «Я сознаю», вряд ли ему удастся обрести сознание…

Ну а Бострому и компании хорошо бы иметь побольше уважения к свершениям эволюции. Трансгуманисты парадоксальным образом возвращаются к своего рода интеллектуальному архаизму, характеризующемуся дуализмом и спиритуализмом. Рэй Курцвейл, американский гуру технофилии, который ради достижения бессмертия собирается загрузить свой разум в электронный чип, считает, наверное, что мысли парят в воздухе. Он, кстати, предлагает мысленный эксперимент: что, если во время сна заменить наши нейроны электронными компонентами, производящими идентичное «Я»?[54] Почему бы в таком случае не обменять физическое «Я» на цифровое? Согласимся с Курцвейлом в том, что такое цифровое «Я» может в высказываниях походить на свой образец. Но ему все равно придется довольствовать-

---

[53] *Спиноза Б.* Этика. Часть 2. Теорема 27 // Собрание сочинений в 2-х т. Т. 1. С. 311.

[54] *Kurzweil R.* How to Create a Mind. Viking, 2012.

ся лишь вечным подражанием идеям, происхождение которых может быть только органическим. Идея без аффекта – это призрак самой себя. Дивное бессмертие, похожее на бессмертие камня в космическом пространстве, вечно кружащего по своей орбите… Скорее я предложил бы Курцвейлу больше наслаждаться своим смертным телом, менять его отношения движения и покоя[55], чтобы производить новые концепты. Поскольку, как напоминает Спиноза в известном пассаже, мы не знаем всего, на что способно тело: «Устройство человеческого тела по своей художественности далеко превосходит все, что только было создано человеческим искусством»[56]. Это верно как в эпоху первых вычислительных машин, так и во времена ИИ.

Бостром и компания презирают философию. Дело в том, что они не понимают тела.

По всем этим причинам, обоснованным не столько метафизикой разума, сколько науками о жизни, нет никакого риска, что ИИ выйдет из своей коробки и начнет преследовать какие-то экстравагантные цели. Тем не менее траектория развития сверхинтеллекта как гипотеза остается возможной. Однако сначала нужно было бы искусственно произвести гомеостаз, а потому и создать сверхорганизм, способный порождать чувства неизвестной силы и сознающий

---

[55] Здесь я воспроизвожу делёзовскую интерпретацию, представленную, в частности, в его работе «Спиноза: практическая философия».

[56] *Спиноза Б.* Этика. Часть 3. Теорема 2. С. 339.

собственное единство. Сегодня эта траектория «воплощенного сознания» стала предметом повышенного интереса в среде специалистов по компьютерным наукам: интеллекту нужно тело. ИИ, натренированный на искусственном «теле», должен оказаться более способным адаптироваться к неожиданным переменам в ситуации[57]. В отсутствие сверхорганизма IBM предлагает начать с простой картонной коробки, снабженной сенсорами[58]. Потом можно будет представить все ступени на лестнице киборга, но в конечном счете вопрос сверхинтеллекта сводится к проблеме создания жизни. Сможем ли мы, осмелимся ли прийти на смену биологической эволюции, чтобы изобрести преемника *Homo sapiens*? Это открытый вопрос, исследуемый сегодня многими учеными, но он выходит за пределы этой книги. Искусственный интеллект не может обладать сознанием. Однако возможно, что однажды искусственное существо произведет сознательный интеллект…

[57] См.: *Hoffmann M., Pfeifer R.* The Implications of Embodiment for Behavior and Cognition: Animal and Robotic Case Studies // Implications of Embodiment / ed. by W. Tschacher, C. Bergomi. Imprint Academic, 2012.

[58] URL: https://www.ibmbigdatahub.com/blog/embodied-cognitionfuture-ai.

Здравый смысл – самая редкая вещь в мире (среди роботов)

Этот экскурс в биологию позволяет нам увидеть внутренние ограничения ИИ, которые постепенно, при подготовке этой главы, стали мне ясны, хотя вначале я их совершенно не понимал. Дело в том, что машина лишена здравого смысла и юмора – и останется лишенной их в будущем. Очень жаль, поскольку это, возможно, два наиболее ценных качества любого собеседника.

Именно Ян Лекун впервые привлек мое внимание к этому моменту: у ИИ нет здравого смысла, то есть он не располагает общим представлением о мире, независимым от конкретной задачи и позволяющим справляться с неопределенностью, новизной или непредвиденными обстоятельствами. Отсюда и происходит ряд инцидентов, которые кажутся нам случайными ошибками. Отдельные трагикомические иллюстрации предоставляет Google Maps: пешеходы, которым предлагают перейти автобан, или туристы, которых отправляют в австралийскую пустыню. ИИ способен избежать тысячи человеческих ошибок, но при этом может совершать ошибки, немыслимые для человека. «Нельзя переходить автотрассу», – вот что сразу говорит нам здравый смысл, тогда как ИИ должен этому учиться. Если понятие автотрассы не входит в параметры его обучения, ему совершенно незачем

возражать против того, чтобы пешеход выходил на проезжую часть. Как только ситуация отклоняется от нормы и требует применения трансверсального суждения, ИИ теряется.

Описать здравый смысл очень сложно, но он проходит через всю историю философии. Аристотель в своей работе «О душе» представил его в качестве такой формы различения, которая находится как бы на полпути между чувствами и рациональностью (определив его как «общее чувство», *koine aisthesis*). Спустя два тысячелетия Делёз определил его в качестве способности идентификации, соотносящей многообразное с единством, единством мира или Эго[59]. В самом деле, здравый смысл задает нормы суждения, которые являются одновременно расплывчатыми и твердыми. Он позволяет нам уверенно обживать один и тот же универсум; тот, «у кого нет здравого смысла» (что само по себе может служить порицанием), внушает беспокойство, он изгоняется из сообщества. Философы осознают всю важность здравого смысла, который, по Аристотелю, необходим для работы разума, а по Делёзу – для возникновения смысла. Однако философам трудно его охарактеризовать. Отсутствие определения здравого смысла – это как раз то, что причиняет столько мучений ИИ и его программистам…

Чтобы лучше понять это, мне кажется полезным дать слово Декарту, известному среди французских бакалавров тем, что он реабилитировал здравый смысл. Так как я не смог на-

[59] *Делёз Ж.* Логика смысла. Серия 12. «О парадоксе».

нести ему визит, предлагаю вообразить, как могла бы произойти наша встреча[60].

После короткого перелета из аэропорта Лондона я прибываю в Амстердам, на омытую дождем площадь между каналом и протестантской церковью. Туристы выстраиваются дисциплинированными цепочками перед домом Анны Франк, где память о холокосте превратилась в скорбный бизнес со льготными тарифами и аудиогидами. Я иду к дому номер шесть, в нескольких шагах от музея. Не без труда нахожу черную дверь, которая никого не интересует. Я стою перед высоким и узким кирпичным зданием, весьма скромным по голландским критериям, увенчанным странной трубой в виде котелка. Вопреки показухе этого торгового города, шторы на окнах опущены. Во внешнем облике дома – ничего особо привлекательного.

Я позвонил, дверь открылась сама собой. Если бы не мои контакты с францисканцами, я никогда не добился бы этой встречи: господин Декарт жил уединенно и постоянно менял адрес. Когда я поднимался по темной и крутой лестнице, у меня сосало под ложечкой. Тяжелый запах подтухшего мяса исходил, казалось, от самих стен; вероятно, это следствие анатомических рассечений. Хотя я и повторял про себя вслед за Паскалем и Жаном-Франсуа Равелем, что Декарт – «неуверенный и бесполезный», это мне не слишком помо-

[60] Эта встреча основана на «Рассуждении о методе», «Трактате о человеке» и «Страстях души».

гало. Слуга, появившийся неведомо откуда, без единого слова проводил меня в комнату на верхнем этаже, где меня ожидал философ, греющий ноги на керамической печке. Он сидел в тени, в своем вечном черном костюме. Его лицо над пышным воротником, казалось, парило в комнате само по себе. Лицо скорее военного, чем античного мудреца, крючковатый нос и утомленный взгляд. Он моего возраста или на несколько лет младше, но мне показался человеком другого поколения, словно бы у меня только-только закончилась молодость, а он уже вступил в старость. Может, я скоро буду на него похож? Определенно, перед моим приходом он долго занимался своим трудом. Я почувствовал некоторую меланхолию и зависть. В качестве утешения, пусть мелочного и смешного, я позволил себе обратить внимание на то, что его бородка плохо подстрижена. Что за небрежность! Пол под моими ногами скрипел, неприятно нарушая тишину.

– Спасибо, что приняли меня, – начал я по привычке к долгим приветствиям, которую приобрел в Кремниевой долине.

Он резко меня оборвал:

– Хватит болтать.

Другого стула в комнате не было. Рене, казалось, это не волновало, поэтому я решил стоять и вынул из сумки записную книжку, чтобы делать заметки.

– Итак, учитель, это по поводу здравого смысла…

– Самой распространенной вещи на свете! – пригвоздил

он.

Сделав вид, что записываю, я изобразил сосредоточенное выражение лица. Мне пришлось проделать весь этот путь не для того, чтобы выслушивать банальности. Подняв ручку, я обратил на Рене взгляд, полный надежды. Но он не стал развивать свою мысль.

– Но почему же? – осмелился спросить я.

Он поднял густые, слегка скошенные брови, которые придавали его лицу какой-то азиатский оттенок.

– Вам нужен порядок причин?

Может, лучше бы я ушел, больше ни о чем не спрашивая. Всегда мог бы сказать, что видел Декарта, процитировать его фразу о «самой распространенной вещи» и этим, собственно, и ограничиться.

– Не следует упорствовать с поиском причин в здравом смысле, – продолжал он, не глядя на меня, словно бы говорил сам с собой. – В первую очередь, это очевидность: без здравого смысла, которым вас одарила природа, как вы могли бы сориентироваться в незнакомом городе, поднимать ноги вровень со ступеньками лестницы в этом доме и найти общие слова, которые позволяют нам вести этот диалог?

Таким образом, эта неловкая беседа была вдруг повышена до ранга диалога. Я, немного надувшись, поддакнул.

– Представим на мгновение, что какой-нибудь злокозненный гений лишил нас здравого смысла: наши органы чувств продолжали бы воспринимать окружающие вещи, осмелюсь

предположить даже, что наш разум мог бы формировать их ясные и отчетливые идеи, но как мы смогли бы схватывать отношения между всеми вещами, образующими наш мир? Может быть, тогда следовало бы обратиться к интеллектуальному познанию, чтобы при помощи одного лишь рассуждения вывести движения мускулов и правила морали? Это невозможно. Мы бы погибли, не сделав и трех шагов.

– Но откуда берется этот здравый смысл? Вы же детерминист…

– Детерминист! – восклицает он. – Что еще за болезнь такая?

Действительно, это слово тогда еще не придумали. Я закусил губу. Неподходящий момент для оплошности.

– Я хочу сказать, что, поскольку наше тело напоминает часы, сделанные из колесиков и пружин…

– Ха! Вы очень милы.

Рене, похоже, наконец оживился. Ничто не возбуждало его больше перспективы опровержения. Особенно когда это опровержение его собственных сочинений.

– Необходимо, чтобы общий смысл…

– Вы имеете в виду здравый смысл?

– Это одно и то же. Перестаньте меня перебивать. Необходимо, чтобы здравый смысл размещался где-то между разумом и механикой тела. Я вам покажу, где именно.

Он оборачивается ко мне, откидывает рукой прядь длинных волос и стучит сзади по своему черепу.

– Шишковидная железа! Единственная часть нашего мозга, которая не является парной, именно она обеспечивает единство тела и души.

Мне так и хочется вздохнуть. Ох уж эта шишковидная железа… Одно из самых неудачных изобретений в истории философии.

– Именно эта небольшая железа объединяет все наши образы и позволяет разуму оказывать действие на нервы. Природа или привычка связали каждое движение железы с определенной мыслью, чтобы можно было давать телу инструкции. Так, когда мы говорим, то не думаем о том, как движутся губы и язык, поскольку простая идея слов воздействует на шишковидную железу, которая сама уже действует на мускулы посредством животных д*у*хов. То же самое происходит и с самыми сложными вопросами морали. Мы следуем инстинктивным правилам, не нуждаясь в их полной формулировке.

Я не смог удержаться от иронии.

– Но как развить здравый смысл? Может, надо щекотать шишковидную железу?

– Даже те, кого всего труднее удовлетворить в каком-либо другом отношении, обыкновенно не стремятся иметь здравого смысла больше, чем у них есть.

На этом высказывании Рене прикрыл глаза, приняв то выражение, которое можно увидеть на его профиле в Facebook. Возможно, наступил час его медитации. Я смиренно покло-

нился и, прыгая через несколько ступенек, спустился по все так же воняющей лестнице. Какое это было облегчение – выйти на свежий воздух! Четверть часа с Декартом – и вот я уже начал во всем сомневаться.

Прогуливаясь вдоль каналов, я попытался вернуться к нити, которая вела от Декарта к Яну Легуну и от Амстердама к Нью-Йорку. Декарту нужно было ввести гипотезу шишковидной железы, над которой сейчас, по прошествии времени, так легко посмеяться, поскольку он смутно понимал, что здравый смысл не может сводиться к простому интеллектуальному рассуждению, к процессу обработки данных. Человеческий разум не просто калькулятор, а тело не часы; и наоборот, сочетание ИИ и андроида никогда не сможет заменить собой способность к суждению, имеющуюся у человека из плоти и крови. Шишковидная железа не допускает чисто механистического подхода к биологии. Здравый смысл – вот что отличает мозг от компьютера.

Конец ознакомительного фрагмента.