

АЛГОРИТМ СУДНОГО ДНЯ

КАК **FACEBOOK**,
GOOGLE,
MICROSOFT
APPLE И ДРУГИЕ
КОРПОРАЦИИ СОЗДАЮТ
ИСКУССТВЕННЫЙ
СУПЕРИНТЕЛЛЕКТ
И ПОЧЕМУ
ЭТО ПРИВЕДЕТ
К КАТАСТРОФЕ

Э М И У Э Б Б

 **БОМБОРА**
ИЗДАТЕЛЬСТВО

БукТех. Книги про технологии

Эми Уэбб

**Алгоритм судного дня. Как
Facebook, Google, Microsoft, Apple
и другие корпорации создают
искусственный суперинтеллект и
почему это приведет к катастрофе**

«ЭКСМО»

2019

УДК 004.8
ББК 32.813

Уэбб Э.

Алгоритм судного дня. Как Facebook, Google, Microsoft, Apple и другие корпорации создают искусственный суперинтеллект и почему это приведет к катастрофе / Э. Уэбб — «Эксмо», 2019 — (БукТех. Книги про технологии)

ISBN 978-5-04-166373-5

Искусственный интеллект, несмотря на название, творение рук человеческих. А ИИ, разработанный крупнейшими международными корпорациями, такими как Google и Facebook, — это еще и мощнейший инструмент воздействия на людей и общество. Знаете ли вы, как он работает и что лежит в его основе? А задумываетесь ли вы, что будет, если этот всемогущий ИИ кто-то пожелает использовать во зло? Книга о будущем человечества в эпоху глобального корпоративного супер-ИИ от генерального директора Future Today. В формате PDF А4 сохранен издательский макет книги.

УДК 004.8

ББК 32.813

ISBN 978-5-04-166373-5

© Уэбб Э., 2019

© Эксмо, 2019

Содержание

Введение	6
Часть 1	12
Глава 1	12
Есть ли разум внутри машины?	14
Может быть, ум и машина просто следуют алгоритму?	17
Можно ли построить мыслящую машину?	18
Мыслит ли мыслящая машина?	20
Лето и зима ИИ	22
Конец ознакомительного фрагмента.	25

Эми Уэбб
Как Facebook, Google, Microsoft,
Apple и другие корпорации создают
искусственный суперинтеллект и
почему это приведет к катастрофе

Моему отцу, Дону Уэббу, умнейшему из известных мне людей

© Вантух К.А., перевод на русский язык, 2022

© Оформление. ООО «Издательство «Эксмо», 2022

Введение

Пока не поздно

Искусственный интеллект уже среди нас, но выглядит при этом совсем не так, как мы предполагали. Он незаметен – и в то же время лежит в основе финансовых систем, электрических сетей и цепочек поставок розничной торговли. Это его невидимая рука указывает нам дорогу, когда мы ведем машину, решает за нас, какое слово мы собирались набрать на клавиатуре, подсказывает, что нам следует покупать, смотреть, слушать, читать. Именно на этой технологии сейчас строится наше будущее, поскольку она проникает во все аспекты нашей жизни: здравоохранение, жилищное строительство, сельское хозяйство, транспорт, спорт и даже в любовь, секс и смерть.

Искусственный интеллект (ИИ) – не очередная тенденция в среде высоких технологий, не модное словечко и не временное увлечение. Это третья эпоха в развитии вычислительной техники. Мы оказались свидетелями радикальных перемен и в этом смысле похожи на современников промышленной революции. Вначале никто не замечал никаких изменений, ведь в сравнении с продолжительностью их жизни они происходили постепенно. Когда же все кончилось, мир уже выглядел иначе: главными сверхдержавами стали Великобритания и Соединенные Штаты, а их индустриальная, военная и политическая мощь оказалась достаточной, чтобы задать облик следующего столетия.

Сейчас все и повсюду бесконечно обсуждают ИИ и его значение для будущего каждого из нас. Обычные темы вам известны: роботы отберут у нас рабочие места, роботы перевернут экономику с ног на голову, роботы в конце концов станут убивать людей. Замените слово «машина» словом «робот», и мы вернемся к дебатам, которые велись 200 лет назад. Естественно размышлять о последствиях новых технологий для наших рабочих мест и нашей способности зарабатывать деньги, поскольку мы уже видели революционные перемены в очень большом количестве отраслей. Понятно, что, думая об ИИ, мы неизбежно вспоминаем компьютер HAL 9000 из фильма «Космическая одиссея 2001 года», WOPR из «Военных игр», Скайнет из «Терминатора», Роза из «Джетсонов», Делорес из Westworld или любые другие антропоморфные воплощения ИИ из популярной культуры. Тем, кто не занят непосредственно в сфере ИИ, будущее представляется фантастическим или пугающим, причем в обоих случаях – в силу неверных причин.

Те, кто не участвует в каждодневной работе над ИИ, не способны корректно считывать контуры будущего, поэтому когда обычная публика пытается обсуждать ИИ, то либо говорит о роботах-завоевателях, которых видела в кино, либо относится к теме с маниакальным безудержным оптимизмом. Подобное отсутствие промежуточных тонов – своего рода проблема разработки ИИ: одни резко преувеличивают пользу от него, другие видят в нем непобедимое оружие.

Я это знаю, поскольку большую часть прошлого десятилетия посвятила исследованию ИИ и встречаю с людьми и организациями как внутри, так и вне этой экосистемы. Я выступала в качестве консультанта многих компаний, находящихся в самом центре работ по ИИ, включая Microsoft и IBM. Я встречалась с людьми, которые не входят в подобные фирмы, но обладают существенными финансовыми интересами в них: венчурными капиталистами, управляющими частными инвестиционными фондами, высокопоставленными должностными лицами в Министерстве обороны и Государственном департаменте, различными законодателями, считающими, что единственная перспектива отрасли – это регулирование со стороны государства. Кроме того, у меня были сотни встреч с академическими учеными и инженерами, работающими «в поле». Среди людей, непосредственно занятых в области ИИ, почти не встре-

тишь ни крайних апокалиптических, ни утопических взглядов на будущее, которые нам, как правило, преподносят в новостях.

Причина заключается в том, что, подобно ученым, работающим в любой области науки, те, кто действительно создают будущее ИИ, стараются сдерживать ожидания. Выдающиеся, этапные достижения требуют терпения, времени, денег и упорства – а мы имеем привычку об этом забывать. Они настойчиво, шаг за шагом, работают над задачами невероятной сложности, зачастую практически не продвигаясь вперед. Эти люди обладают живым умом, хорошо знают жизнь и, по моему опыту, сочувственны и внимательны.

Практически все они работают в девяти технологических гигантах, создающих ИИ ради лучшего будущего для всех нас: это Google, Amazon, Apple, IBM, Microsoft и Facebook в Соединенных Штатах и Baidu, Alibaba и Tencent в Китае. Я твердо верю, что руководители этих девяти компаний руководствуются глубоким чувством альтруизма и желанием служить высшему благу; они ясно видят, как ИИ может улучшить здоровье и долголетие, решить надвигающиеся на нас проблемы с климатом, вывести миллионы людей из бедности. Мы уже видим вполне заметные положительные результаты их работы во всех отраслях человеческой деятельности и повседневной жизни.

Проблема заключается в том, что внешние силы, оказывающие давление на девять технологических гигантов – и, следовательно, на всех работающих внутри экосистемы, – имеют свои планы, идущие вразрез с их вполне пристойными намерениями относительно нашего будущего. Обвинить можно многих и во многом.

В Соединенных Штатах безжалостные требования рынка и нереальные ожидания, предъявляемые к новым продуктам, сделали невозможным долгосрочное планирование. Мы ждем от Google, Amazon, Apple, Facebook, Microsoft и IBM объявлений о новаторских продуктах в области ИИ на каждой ежегодной конференции, как будто прорывы в области инженерии могут происходить по расписанию. Если какая-либо из этих компаний не предложит нам более впечатляющего продукта, чем в прошлом году, мы решим, что они потеряли форму. Или спросим себя, не прошла ли эпоха ИИ. Или усомнимся в компетентности их руководства. Ни разу мы не давали им возможности спокойно поработать несколько лет, не требуя удивлять нас через регулярные промежутки времени. Упаси бог, какая-нибудь компания из списка решит не делать официальных объявлений несколько месяцев – мы решим, что молчание указывает на наличие тайного проекта, разумеется, способного только огорчить нас.

У правительства Соединенных Штатов нет генеральной стратегии как в отношении ИИ, так и в отношении нашего будущего на отдаленную перспективу. Поэтому вместо создания скоординированной общенациональной стратегии – со специальной организационной структурой внутри правительства, укреплением международных альянсов, подготовкой наших вооруженных сил к будущему характеру военных действий – Соединенные Штаты просто выставили ИИ за порог политики. Вместо того чтобы финансировать основные исследования по ИИ, федеральные власти фактически передали научно-исследовательскую работу в этой области на аутсорсинг коммерческому сектору и Уолл-стрит. Вместо того чтобы видеть в ИИ возможность создавать новые рабочие места и стимулировать экономический рост, американские законодатели обращают внимание только на повсеместную безработицу в технологической отрасли. В свою очередь законодатели склонны выдвигать всевозможные обвинения в адрес американских технологических гигантов, в то время как могли бы пригласить их к участию в принятии решений на самых высоких уровнях стратегического планирования в правительстве (в их теперешнем виде). У наших пионеров ИИ нет выбора, кроме как непрерывно конкурировать друг с другом за непосредственную, доверительную связь с вами, со мной, с нашими школами, больницами, городами, коммерческими предприятиями.

Наша проблема в Соединенных Штатах заключается в трагическом неумении заглядывать в будущее. Мы сосредоточены на сиюминутном, планируем не больше, чем на несколько

лет вперед. Подобное мышление предпочитает сиюминутные же технологические достижения, но избавляет нас от необходимости думать, какими путями может пойти развитие технологии, как предсказывать результаты и последствия наших действий больше чем на один ход. Мы слишком легко забываем о том, что наши сегодняшние действия могут иметь серьезные последствия. Неудивительно, что мы по существу отдали строительство нашего будущего на аутсорсинг шести частным компаниям, чьи ценные бумаги свободно обращаются на бирже, чьи достижения впечатляют, но чьи финансовые интересы далеко не всегда совпадают с тем, что было бы предпочтительно для наших личных свобод, сообществ и демократических идеалов.

В то же время направление разработки ИИ в Китае диктуется грандиозными амбициями государства. Китай спешно закладывает фундамент, чтобы в будущем превратиться в бесспорного мирового гегемона в области ИИ. Китайское руководство представило свой план разработки ИИ следующего поколения, цель которого – добиться мирового лидерства в ИИ к 2030 году со стоимостью отрасли внутри страны как минимум 150 млрд долларов¹. Согласно этому плану, часть фонда национального благосостояния направляется на создание новых лабораторий и стартапов, а также учебных заведений, которые должны будут вырастить следующее поколение талантов именно для ИИ². В октябре того же года президент Китая Си Цзиньпин объявил о своих планах в отношении ИИ и больших данных в подробной речи перед тысячами партийных функционеров. ИИ, сказал он, поможет Китаю превратиться в одну из самых передовых экономик мира. Уже сейчас объем китайской экономики в 30 раз больше, чем был три десятилетия назад. Baidu, Tencent и Alibaba, будучи гигантскими компаниями, чьи бумаги обращаются на бирже, все равно обязаны подчиняться воле Пекина, что типично для крупного китайского бизнеса.

Огромное население Китая – 1,4 млрд человек – означает, что в его распоряжении находится самый богатый и, возможно, самый важный природный ресурс эпохи ИИ: данные о людях. Чтобы добиться высокой точности от алгоритма распознавания образов, требуются огромные объемы данных, и поэтому китайские системы распознавания лиц, например Megvii или SenseTime, так привлекательны для инвесторов. Все данные, которые китайские граждане генерируют, когда звонят друг другу, делают покупки онлайн или публикуют фото в социальных сетях, помогают Baidu, Alibaba и Tencent создавать лучшие в своем классе системы ИИ. Важное преимущество Китая – отсутствие ограничений, связанных с неприкосновенностью частной жизни и безопасностью, которые отчасти могут тормозить прогресс в Соединенных Штатах.

Нам следует рассматривать планы разработки ИИ в более широком контексте грандиозных планов Китая. В апреле 2018 года президент Си выступил с большой речью, где поставил задачу превратить Китай в глобальную компьютерную сверхдержаву. Государственное информационное агентство Синьхуа воспроизвело отрывки из нее, где описана сеть для администрирования киберпространства и Интернет, который «распространял бы положительную информацию, держался бы верной политической линии, направлял бы общественное мнение и ценности в правильную сторону»³. Авторитарный закон, по которому Китай хочет заставить жить всех нас, не имеет ничего общего со свободой слова, рыночной экономикой и распределенным управлением – ценностями, которыми мы дорожим на Западе.

ИИ упоминается в ряде законодательных актов Китая, которые нацелены на тотальный контроль над всей информацией, создаваемой внутри границ страны, и перехват данных ее

¹ Paul Mozur, “Beijing Wants AI to Be Made in China by 2030,” *New York Times*, July 20, 2017, <https://www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html>. – *Прим. авт.*

² Tom Simonite, “Ex-Google Executive Opens a School for AI, with China’s Help,” *Wired*, April 5, 2018, <https://www.wired.com/story/ex-google-executive-opens-a-school-for-ai-with-chinas-help/>. – *Прим. авт.*

³ “Xinhua Headlines: Xi outlines blueprint to develop China’s strength in cyberspace,” *Xinhua*, April 21, 2018. http://www.xinhuanet.com/english/2018-04/21/c_137127374_2.htm. – *Прим. авт.*

резидентов, а также граждан государств – стратегических партнеров. Один из таких актов требует хранить все данные о китайских гражданах на серверах, расположенных на территории Китая. Тем самым государственные службы получают возможность свободного доступа к этим данным. Другая инициатива, «Облако китайской полиции», разработана для контроля и отслеживания местоположения душевнобольных, критиков режима и уйгуров – этнического меньшинства, исповедующего ислам. В августе 2018 года Организация Объединенных Наций заявила о наличии у нее достоверных сведений, что Китай удерживает миллионы уйгуров в тайных лагерях на дальнем западе страны⁴. Так называемая Китайская интегрированная программа совместных действий использует ИИ для поиска отклонений от общепринятого социального поведения – например несвоевременной оплаты счетов. Система социального кредита, согласно удачной фразе из ее основополагающего документа, разработана ради создания идеального общества, «позволяя достойным доверия свободно перемещаться, в то время как дискредитировавшим себя будет трудно сделать хотя бы шаг⁵». Понятие «достойный доверия» определяется на основании численных показателей, например за «героические поступки» (гражданину начисляются баллы) и штрафы за нарушение правил дорожного движения (баллы вычитаются). Граждане с низким рейтингом могут столкнуться с трудностями при приеме на работу, покупке дома или попытке отдать детей в школу. В некоторых городах портреты граждан с высоким рейтингом выставляются на доске почета⁶. В других городах, таких как Шаньдун⁷, фотографии людей, нарушающих правила перехода улицы, выставляются на цифровых рекламных стендах на всеобщее обозрение и автоматически отправляются в Weibo, популярную социальную сеть⁸. Если все это кажется слишком невероятным, вспомните, что Китай смог успешно провести в жизнь политику одного ребенка, чтобы сдержать рост населения.

Эти проекты и инициативы рождены в узком кругу советников президента Си Цзиньпина, которые примерно десять последних лет сосредоточены именно на превращении Китая в мировую сверхдержаву – и соответствующем изменении восприятия Китая окружающим миром, – а также реформах внутри страны. Сегодня Китай более авторитарен, чем при любом из предыдущих руководителей со времен Мао Цзедуна, и разработка и внедрение ИИ – необходимое условие сохранения такого положения дел. Инициатива «Один пояс и один путь» – на самом деле масштабный глобальный стратегический план, замаскированный под план строительства инфраструктуры, которая повторяет маршруты Шелкового пути, что связывал когда-то Китай с Европой через Ближний Восток и Африку. Китай не просто строит мосты и шоссе – он экспортирует технологию слежения и собирает данные, увеличивая таким образом влияние Коммунистической партии Китая в мире в противовес нашему либеральному демократическому порядку. Инициатива Global Energy Interconnection – еще один национальный стратегический план, который поддерживает президент Си. Он предусматривает создание первой в мире глобальной сети электроснабжения, управляемой из Китая.

Там уже научились масштабировать технологию передачи электроэнергии по сверхвысоковольтным кабелям, что позволяет доставлять электричество в Шанхай с дальнего запада

⁴ Stephanie Nebehay, “U.N. says it has credible reports that China holds million Uighurs in secret camps,” Reuters, August 10, 2018. <https://www.reuters.com/article/us-china-rights-un/u-n-says-it-has-credible-reports-that-china-holds-million-uighurs-in-secret-camps-idUSKBN1KV1SU>. Прим. ред.

⁵ Simina Mistreanu, “Life Inside China’s Social Credit Laboratory,” Foreign Policy, April 3, 2018. <https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/>. – Прим. авт.

⁶ Ibid. – Прим. авт.

⁷ На самом деле Шаньдун – провинция Китая, а в статье, на которую ссылается автор, речь идет о событиях в ее столице, городе Цзинань. – Прим. пер.

⁸ “China Shames Jaywalkers through Facial Recognition,” Phys.org, June 20, 2017, <https://phys.org/news/2017-06-china-shames-jaywalkers-facialrecognition.html>. – Прим. авт. Глава 1

страны – и сейчас Китай заключает договоры, превращающие его в поставщика электроэнергии для соседних стран.

Эти инициативы, как и многие другие, – остроумный способ надолго обеспечить расширение своей «мягкой власти». Это великолепный ход со стороны Си Цзиньпина, за отмену предельного срока пребывания на посту которого КПК проголосовала в марте 2018 года, тем самым позволив ему по сути оставаться президентом пожизненно. Конечная цель стратегии президента Си предельно проста: построить мировой порядок, обеспечивающий Китаю лидерство де-факто. Тем не менее во время расширения сферы дипломатического влияния Китая Соединенные Штаты необъяснимым образом перестали замечать долгосрочные глобальные союзы и соглашения с его участием, когда президент Трамп возвел новый «бамбуковый занавес».

Если говорить о будущем ИИ, то его разработка сейчас движется двумя путями, часто не соответствующими представлениям о благе человечества. Усилия Китая являются частью скоординированного плана построения нового мирового порядка с Си Цзиньпином во главе, а в Америке главными движущими силами служат рыночные факторы и консьюмеризм. Сочетание этих двух факторов представляет собой слабое место, потенциально угрожающее всем нам. Разрешение этой проблемы – самый важный вопрос для будущего развития ИИ, именно в этом и заключается задача настоящей книги. Может быть, компании «Большой девятки» преследуют одну и ту же благородную цель – взломать код машинного интеллекта, чтобы строить системы, способные мыслить подобно человеку, – но со временем результаты их работы могут нанести непоправимый вред человечеству.

В принципе я верю, что ИИ представляет собой положительную силу, которая принесет много пользы следующим поколениям и поможет построить будущее, соответствующее нашим самым идеалистическим представлениям о нем.

Но я прагматик. Мы все знаем, что даже люди, которые руководствуются самыми благими намерениями, способны причинить огромный вред. Говоря о технологии, особенно об ИИ, всегда следует иметь план на случай злоупотребления ею, как случайного, так и намеренного. Особенно важно это сейчас, когда ИИ захватывает все сферы человеческой деятельности: глобальную экономику, рынок труда, сельское хозяйство, транспорт, банковскую сферу, контроль за состоянием окружающей среды, образование, вооруженные силы и национальную безопасность. Вот почему, если нынешние направления развития ИИ в Соединенных Штатах и Китае сохранятся, 2069 год может оказаться совсем не похожим на 2019-й. По мере того как структуры и системы, управляющие обществом, постепенно перейдут на использование ИИ, мы увидим, что принятые ими от нашего имени решения выглядят совершенно разумными только с точки зрения машины, а не с нашей.

Мы, люди, быстро теряем бдительность – как раз тогда, когда машины пробуждаются. Мы уже прошли несколько важных этапов технологического и геополитического развития ИИ, но при этом с каждым значительным шагом вперед ИИ становится все менее заметен нам. Наши данные анализируются и уточняются все менее очевидным для нас образом, а наша способность понимать, каким образом автономные системы принимают решения, снижается. Таким образом, уже сейчас мы не в состоянии понять, каким образом ИИ влияет на нашу жизнь, и этот разрыв будет расти по экспоненте с течением времени, когда для нас минуют годы, а потом и десятилетия. Сократить этот разрыв насколько возможно, критикуя нынешний путь развития ИИ, – вот задача, ради которой я пишу эту книгу. Моя цель – сделать обсуждение искусственного интеллекта более демократичным, снабдить вас дополнительными знаниями относительно того, что нас ждет, а также объяснить, пока еще не поздно, как последствия развития ИИ коснутся именно вас и почему они будут совершенно ощутимыми и конкретными.

Человечество проходит через экзистенциальный кризис в самом буквальном смысле слова, поскольку никто не занимается вопросом, ключевым для ИИ с самого начала работ над

ним: что произойдет с обществом, когда мы передадим власть системе, разработанной небольшой группой людей для того, чтобы она принимала решения, которые касаются всех? Что произойдет, если эти решения склонятся в сторону рыночных факторов или интересов амбициозной политической партии? От ответа зависит, какие именно возможности откроются перед нами в будущем, каким образом нам станут закрывать доступ к тому или иному ресурсу, какие социальные условности возникнут внутри наших обществ, по каким правилам заработает экономика наших стран и каким образом мы окажемся связаны с другими людьми.

Эта книга не об обычных спорах вокруг ИИ. Она – одновременно предостережение и набросок лучшего будущего. Она ставит под вопрос нелюбовь нас, американцев, к долгосрочному планированию, и подчеркивает неготовность к приходу ИИ в сфере бизнеса, образования, государственного управления. Она рисует впечатляющую картину скоординированных между собой геополитических, экономических и дипломатических планов Китая, направленных к единой цели – установлению мирового господства. И она требует героического руководства в исключительно трудных обстоятельствах. Потому что, как вы сами увидите, нашему будущему требуется герой.

Текст, который последует ниже, представляет собой призыв к действию в трех частях. Из первой вы узнаете, что такое ИИ и какую роль в его разработке сыграла «Большая девятка». Мы также в деталях рассмотрим положение, в котором находятся американские члены «Большой девятки» и китайские – Baidu, Alibaba и Tencent.

Вторая часть подробно описывает вероятные варианты будущего, исходя из прогноза развития ИИ на 50 лет. Среди трех сценариев, которые вы увидите, окажется и оптимистический, и прагматический, и катастрофический, и в них будут показаны как возможности, так и риски, с которыми сопряжен переход от слабого к сильному ИИ. Все эти сценарии несут высокую эмоциональную нагрузку. Они построены на основе моделирования с использованием фактических данных и дают вам интуитивное представление о том, в каком направлении может развиваться ИИ и как в результате этого изменится наша жизнь.

В третьей части я предлагаю стратегические и тактические решения всех проблем, перечисленных в сценариях, а также конкретный план исправления настоящего. Третья часть, по моему замыслу, должна заставить нас действовать, поэтому в ней содержатся конкретные рекомендации для правительств, «Большой девятки» и даже вас лично.

* * *

Каждый человек из числа живущих сегодня способен сыграть серьезную роль в будущем ИИ. Решения, которые мы принимаем относительно ИИ сейчас – казалось бы, даже самые незначительные, – навсегда изменят ход человеческой истории. Машины пробуждаются, и в один прекрасный день мы можем понять, что вопреки нашим надеждам и альтруистическим амбициям воздействие наших систем ИИ на человечество было катастрофическим.

Но подобное развитие событий отнюдь не предопределено.

«Большой девятке» в моем рассказе вовсе не отводится роль злодея. Напротив, входящие в нее компании – наша главная надежда на лучшее будущее.

Переверните страницу. Мы не можем пассивно ждать, что принесет нам будущее. ИИ уже среди нас.

Часть 1

Духи внутри машины

Глава 1

Разум и машина: очень краткая история ИИ

Корни современного ИИ берут свое начало в далеком прошлом, задолго до того, как «Большая девятка» создала наших электронных помощников: Сири, Алексу и их китайского коллегу Тян Мао. И за все это время не было создано единого определения ИИ, подобного тому, что существует для других технологий. Когда дело касается ИИ, оказывается, что сложно дать его конкретное описание, поскольку он представляет собой разные вещи, даже когда отрасль продолжает развиваться. То, что вполне можно было назвать ИИ в 1950-х – калькулятор, способный выполнять деление столбиком, – сегодня едва ли воспринимается как шедевр технологии. Это явление известно под названием «парадокса невозможного» – как только новая технология становится частью мейнстрима, мы перестаем ее замечать. Мы больше не воспринимаем ее как ИИ.

В самой базовой форме ИИ представляет собой систему, самостоятельно принимающую решения. Действия, выполняемые ИИ, повторяют работу человеческого разума или подражают ей. Примерами могут служить распознавание звуков и изображений, решение задач, понимание естественного языка, использование стратегии для достижения целей. Некоторые системы ИИ огромны и способны осуществлять миллионы вычислений, в то время как другие – узки и предназначены для решения всего одной задачи, например поиска обценной лексики в сообщениях электронной почты.

В своих рассуждениях мы постоянно возвращаемся к одним и тем же вопросам: могут ли машины думать? Что означает глагол «думать» в применении к машине? Что означает он в применении к нам? Что такое мысль? Как мы можем быть уверены – однозначно, с полной определенностью, – что мы на самом деле думаем свои собственные оригинальные мысли? Эти вопросы живут рядом с нами в течение столетий и обладают основополагающим значением как для истории ИИ, так и для его будущего.

Когда мы задаемся вопросом, каким именно образом мыслят люди и машины, главной проблемой оказывается, что слово «мыслить» неотделимо от слова «ум». Словарь *Merriam-Webster Dictionary* определяет слово «мыслить» как «формировать или иметь в уме», а *Oxford Dictionary*⁹ объясняет, что это означает «активно использовать ум для формирования взаимосвязанных идей». Если мы посмотрим в словаре значение слова «ум», то окажется, что и *Merriam-Webster*, и *Oxford* определяют его в контексте слова «сознание». Но что такое сознание? Согласно обоим словарям, это качество или состояние, при котором мы осознаем действительность и реагируем на предлагаемые ею стимулы¹⁰. Различные группы специалистов: психологи, нейробиологи, философы, теологи, ученые в области этики или информатики – все они будут подходить к определению термина «мышление» со своих позиций.

Если мы ищем столик в любимом ресторане при помощи Алексы, она, как и мы, будет обсуждать еду так, словно осознает действительность и способна реагировать на внешние сти-

⁹ Именно эти два толковых словаря играют основную роль для американского и британского варианта английского языка соответственно. – *Прим. пер.*

¹⁰ Один из вариантов значения слова «мысль» в словаре Ожегова тоже опирается на термин «сознание»: «То, что заполняет сознание, дума». – *Прим. пер.*

мулы. При этом Алекса никогда не чувствовала хрустящего яблока на зубах, пикантных маринованных огурчиков или газированной воды на языке, прикосновения густого арахисового масла к нёбу. Попросите Алексу описать свойства этих продуктов, и ее ответ будет опираться на ваш собственный опыт. У Алексы нет рта – как же она может ощущать вкус пищи, подобно вам?

Вы – уникальная в биологическом смысле личность, и потому слюнные железы и вкусовые сосочки расположены у вас не совсем так, как у меня. Тем не менее мы оба знаем, что такое яблоко, каково оно в общем на вкус, какова его текстура, как оно пахнет. В какой-то момент жизни мы научились узнавать яблоко при помощи методики, которую специалисты называют «обучением с подкреплением»: кто-то объяснил нам, что такое яблоко и чем оно отличается от других фруктов. Потом, со временем и без нашего осознанного участия, наши автономные биологические системы распознавания образов довели до совершенства навык узнавания яблока, даже при недостаточных данных. Увидев черно-белое, двумерное, схематичное к тому же изображение яблока, мы знаем, что это такое, – несмотря на то, что у нас нет сведений о вкусе, запахе, хрусте и других вещах, сигнализирующих мозгу: «Это яблоко». Способы, которыми мы и Алекса узнали что-то о яблоках, куда ближе друг к другу, чем можно было бы подумать.

Алекса вполне компетентна, но вот *разумна* ли она? Должно ли ее машинное восприятие обладать всеми качествами человеческого, чтобы мы приняли ее способ «мышления» как адекватное отражение своего собственного? Американский психолог методов обучения Бенджамин Блум основную часть своей академической карьеры посвятил исследованию и классификации состояний ума. В 1956 году он опубликовал систему, позже известную как «таксономия Блума», описывавшую цели обучения и достижения учащихся в образовательном процессе. Базовый уровень – это запоминание фактов и основных понятий, за ним по порядку следуют понимание идей, применение знаний в новой ситуации, экспериментальный анализ информации и выстраивание логических связей, оценка, защита и определение ценности информации и, наконец, создание оригинальной работы. Будучи маленькими детьми, мы в основном заняты запоминанием и пониманием. Например, нам сначала следует запомнить, что в бутылке находится молоко, а только после этого – что у бутылки есть лицевая и тыльная сторона, даже если мы одной из них не видим.

Компьютеры обучаются, следуя той же иерархии. Например, в 2017 году система ИИ под названием Ampeg создала и записала музыку для альбома I AM AI («Я – ИИ»). Аккорды, инструментовку, партии ударных Ampeg разрабатывала сама, используя в качестве исходных параметров жанр, настроение и длину произведения, и при этом среди ее произведений – про-никновенная баллада Break Free, набравшая более 1,6 млн просмотров на YouTube и ставшая хитом на обычном радио. Прежде чем написать такую песню, Ampeg нужно было узнать качественные характеристики популярной баллады, а также некоторую количественную информацию, например как рассчитывать высоту нот и такты, как распознавать тысячи шаблонов в музыке (аккордовые последовательности, гармонические последовательности, ритмические акценты).

Способность к творчеству, по Блуму, является высшим результатом обучения, но не было ли оно в данном случае просто выученным механическим процессом? Было ли это творчество подобно человеческому? Или его природа была совершенно иной? Думала ли Ampeg о музыке так же, как композитор-человек? Можно было бы попытаться доказать, что «мозг» Ampeg – нейронная сеть, которая использует алгоритмы и данные, ограниченные своим контейнером, – не так уж сильно отличается от мозга Бетховена, который построен из органических нейронов, использует данные, распознает образы и заключен в контейнер его головы. Так ли сильно отличался творческий процесс Ampeg от бетховенского, когда тот создавал свою Пятую симфонию с ее знаменитым началом «та-та-та-ТАМ» и последующим переходом из мажора в минор? За

этими четырьмя нотами следуют гармоническая последовательность, фрагменты гамм, арпеджио и другие ингредиенты, из которых состоит любая композиция. Прислушайтесь внимательно к скерцо в финале, и вы услышите очевидные заимствования из Сороковой симфонии Моцарта, написанной на двадцать лет раньше, в 1788 году. Моцарт испытывал влияние Антонио Сальери и своего друга Франца Йозефа Гайдна, которые, в свою очередь, находились под влиянием своих предшественников, например Иоганна Себастьяна Баха, Антонио Вивальди и Генри Перселла, творивших с середины XVII до середины XVIII века. В их музыке можно услышать следы работ еще более ранних композиторов, таких как Якоб Аркадельт, Жан Мутон и Йоханнес Окегем. Те, в свою очередь, испытывали влияние композиторов самого раннего Средневековья – и цепочку можно проследить дальше, до самой первой записанной композиции, так называемой Сейкилосской эпитафии, выбитой на мраморном столбе, который был обнаружен на территории современной Турции и когда-то служил надгробным памятником. Можно углубляться и дальше в прошлое, к тем временам, когда были изготовлены первые примитивные флейты из костей животных и слоновой кости – то есть примерно 43 000 лет назад. Ученые полагают, что и в более ранние времена наши предки пели, прежде чем начать говорить¹¹.

Человеческое мышление представляет собой результат миллионов лет эволюции. Подобным же образом мышление современного ИИ опирается на долгий эволюционный процесс, начинающийся со времен математиков, философов и ученых древности. Хотя может показаться, что человечество и царство машин всегда шли разными путями, на самом деле наше развитие было взаимосвязано. Homo sapiens извлекал уроки из окружающей его среды, передавал нужные привычки последующим поколениям, приобретал все большее разнообразие и размножался благодаря изобретению совершенных технологий, таких как сельское хозяйство, охотничьи орудия, пенициллин. Понадобилось 11 000 лет, чтобы население Земли увеличилось с шести миллионов человек в неолите до семи миллиардов в наши дни¹².

Представляет ли Алекса себе яблоко так же, как мы, или в самом ли деле «оригинальна» музыка Ampeg, – все это вопрос о том, как мы мыслим себе процесс мышления.

Современный ИИ представляет собой сплав тысячелетнего опыта философов, математиков, ученых, специалистов по робототехнике, художников и теологов. Их задача – как и наша в настоящей главе – заключается в том, чтобы понять связь между мышлением и контейнером для мышления, между человеческим разумом и машинами. Как возникает эта связь благодаря – или *вопреки* – машинам, которые строит «Большая девятка» в Китае и Соединенных Штатах?

Есть ли разум внутри машины?

Основные идеи ИИ ведут свое происхождение из Древней Греции и восходят к истокам философии, логики и математики. Во многих текстах Платона Сократ говорит: «Познай самого себя», – подразумевая, что стать лучше и принимать правильные решения можно, только узнав сначала собственный характер. Аристотель, помимо прочего, изобрел силлогическую логику и нашу первую формальную систему рассуждений методом дедукции. Приблизительно в то же время математик Евклид придумал, как найти наибольший общий делитель двух чисел, и таким образом разработал первый алгоритм. Благодаря этим древнегреческим мыслителям возникли две новые важные идеи: некоторая физическая система может действовать как набор логических правил, и человеческое мышление как таковое может быть символической системой. Вместе с ними возникли и вопросы, которые философы, теологи, ученые задают себе

¹¹ “The Seikilos Epitaph: The Oldest Song in the World,” Wired, October 29, 2009, <https://www.wired.com/2009/10/the-seikilos-epitaph>.

¹² “Population Clock: World,” Census.gov, 2018, <https://www.census.gov/popclock/world>.

уже сотни лет. Не является ли тело сложной машиной? Целое, составленное из тысяч других систем, работающих совместно, подобное старым часам с маятником? А что тогда с разумом? Он тоже сложная машина? Или нечто совершенно иное? Доказать его божественное происхождение или связь между разумом и физическим миром невозможно – как и опровергнуть любую из этих гипотез.

В 1560 году испанский часовщик Хуанело Турриано создал небольшую фигурку механического монаха в качестве дара церкви от лица испанского короля Филиппа II, сын которого чудесным образом исцелился после травмы головы¹³. Монах умел делать удивительные вещи: ходил по столу, поднимал крест и четки, бил себя в грудь в знак покаяния и безмолвно шевелил губами в молитве. Это был первый *автоматон* – механическое подобие живого существа. Слова «робот» еще не существовало, но механический монах был изумительным изобретением, которое наверняка потрясло и пугало зрителей. Вероятно, никому не приходило в голову, что маленький автоматон когда-нибудь сможет не только повторять основные человеческие движения, но заменять человека в заводских цехах, научных лабораториях и кухонных разговорах. Маленький монах вдохновил первое поколение робототехников, чья задача заключалась в том, чтобы создать машины более сложные, чем подражающие человеческим движениям: автоматы вскоре научились писать, танцевать, рисовать. В результате некоторые философы стали задумываться над вопросом, что значит быть человеком. Если можно построить автомат, подражающий поведению человека, не являются ли люди автоматами, созданными божеством? Или мы представляем собой сложные системы, способные к рассуждению и созданию оригинальных идей?

Английский политический философ Томас Гоббс описывал человеческое рассуждение как вычисление в трактате «О теле» (*De Corpore*), входящем в состав его великой трилогии, посвященной естественным наукам, психологии и политике. В 1655 году он писал: «Через рассуждения я понимаю вычисление. А вычислять означает находить сумму многих вещей, складываемых одновременно, или знать разность, когда одна вещь отнимается от другой. Рассуждать, таким образом, то же самое, что складывать или вычитать»¹⁴. Но как нам узнать, обладаем ли мы при этом свободной волей?

В то время, когда Гоббс работал над первой частью своей трилогии, французский философ Рене Декарт опубликовал книгу «Размышления о первой философии», в которой задавался вопросом, как мы можем знать наверняка, что данный нам в ощущениях мир реален. Как можем мы проверить собственное сознание? Какое доказательство могло бы убедить нас, что наши мысли действительно принадлежат нам, а мир вокруг нас реален? Декарт был рационалистом, верившим, что факты могут быть получены путем дедукции. Он предложил знаменитый мысленный эксперимент. Он предложил читателям представить себе демона, якобы создающего иллюзию окружающего нас мира. Если бы физический, чувственный опыт читательницы, сообщающей ей, что она плавает в озере, был бы всего лишь наведен демоном, она не могла бы действительно *знать*, что плавает. Но с точки зрения Декарта, если читатель осознает свое существование, значит, он отвечает критериям знания. «Всякий раз, как я произношу слова *Я есмь, я существую* или воспринимаю это изречение умом, оно по необходимости будет истинным»¹⁵, – писал он¹⁶. Другими словами, факт нашего существования не подлежит сомнению,

¹³ Elizabeth King, “Clockwork Prayer: A Sixteenth-Century Mechanical Monk,” Blackbird 1, no. 1 (Spring 2002), https://blackbird.vcu.edu/v1n1/nonfiction/king_e/prayer_introduction.htm.

¹⁴ “AMA: We Are the Google Brain Team. We’d Love to Answer Your Questions about Machine Learning,” Reddit, August 4, 2016, https://www.reddit.com/r/MachineLearning/comments/4w6tsv/ama_we_are_the_google_brain_team_wed_love_to/.

¹⁵ Декарт Р. Сочинения в 2 томах. Т. 1. М., Мысль, 1989. Цит. по: https://nibiryukov.mgimo.ru/nb_russian/nbr_teaching/nbr_teach_library/nbr_library_classics/nbr_classics_descartes_meditationes_de_prima_philosophia.htm. Пример с читательницей, плавающей в озере, добавлен автором. – Прим. пер.

¹⁶ René Descartes, Meditations on First Philosophy, Second Meditation § 25, 1641, University of Connecticut, <http://selfpace.uconn.edu/class/percep/DescartesMeditations.pdf>.

даже если между нами и реальностью находится вводящий нас в заблуждение демон. Или же «я мыслю, следовательно, существую».

Позже в своем «Трактате о человеке» Декарт утверждал, что люди, вероятно, могли бы изготовить автомат в виде небольшого животного, который был бы неотличим от своего прототипа. Но даже если когда-нибудь нам удалось бы создать механического человека, его никогда нельзя будет спутать с настоящим, утверждал Декарт, поскольку у него не будет разума и, следовательно, души. В отличие от людей, машина никогда не сможет соответствовать критериям знания – она никогда не сможет осознавать себя, как это делаем мы. Для Декарта сознание находилось внутри нас – душа была духом, оживляющим машину нашего тела¹⁷.

Несколькими десятилетиями позже немецкий математик и философ Готфрид Вильгельм фон Лейбниц рассматривал идею, что человеческая душа сама по себе запрограммирована, и утверждал, что при этом человеческий ум оказывается контейнером. Бог создал душу и тело так, чтобы они естественным образом сосуществовали в гармонии. Тело, разумеется, сложная машина, но оно едино с набором божественных инструкций. Наши руки движутся, когда мы захотим, но не мы изобрели механизмы, делающие движение возможным. Если мы осознаем боль или удовольствие, эти ощущения – результат работы заранее запрограммированной системы, непрерывной связи между разумом и телом.

Лейбниц предложил собственный мысленный эксперимент, иллюстрирующий его положение, что мысль и чувства неотделимы от человеческой природы. Представьте себе, что вы заходите на мельницу. Здание – это контейнер, заключающий в себе машины, сырье и рабочих. Это сложная система, составные части которой согласно работают ради общей цели, но разума она иметь не может. «Все, что мы найдем там, – это шестерни и рычаги, приводящие в движение друг друга, но ничего способного отвечать за чувства, – писал он, – Таким образом, чувства нужно искать в простых сущностях, но не в составных, таких, как машины». Его идея заключалась в том, как бы сложны ни были мельница, механизмы или автоматы, человек не может создать машину, способную мыслить или чувствовать¹⁸.

Тем не менее Лейбница увлекала мысль о воспроизведении отдельных аспектов мышления. Несколькими десятилетиями ранее малоизвестный английский литератор Ричард Брейтуэйт, написавший несколько книг о поведении в обществе, мимоходом упоминает «компьютеров» – прошедших сложное обучение людей, способных быстро и точно производить вычисления¹⁹. В то же время французский математик и изобретатель Блез Паскаль, заложивший основы того, что сегодня мы знаем как теорию вероятности, занимался автоматизацией вычислительных задач. Паскаль наблюдал, как его отец кропотливо высчитывает налоги вручную, и хотел упростить ему работу. Поэтому он начал работать над автоматическим вычислителем с зубчатыми колесами и подвижными круговыми шкалами²⁰. Вычислитель работал, и это заставило Лейбница уточнить свою мысль: у машин никогда не будет души, но когда-нибудь удастся построить машину, способную к логическому мышлению на уровне человека. В 1673 году Лейбниц описал «пошаговую машину выводов», новый тип вычислительного устройства, способный принимать решения с использованием двоичной системы²¹. Машина

¹⁷ René Descartes, *Treatise of Man*, trans. T. S. Hall (Cambridge, MA: Harvard University Press, 1972).

¹⁸ Gottfried Wilhelm Leibniz, *The Monadology*, trans. Robert Latta, (1898), <https://www.plato-philosophy.org/wp-content/uploads/2016/07/TheMonadology-1714-by-Gottfried-Wilhelm-LEIBNIZ-1646-1716.pdf>.

¹⁹ Полагают, что слово «компьютер» впервые употреблено в книге *The Yong Mans Gleanings*, написанной Ричардом Брейтуэйтом в 1613 году. В те времена «компьютерами» называли людей, производящих вычисления.

²⁰ “Blaise Pascal,” *Biography.com*, <https://www.biography.com/people/blaise-pascal-9434176>.

²¹ Лейбниц пишет в *De progressionem dyadica*: «[Двоичное] исчисление может быть реализовано при помощи машины, оснащенной отверстиями таким образом, чтобы их можно было открывать и закрывать. Их следует открывать в местах, соответствующих 1, и оставлять закрытыми – в соответствующих 0. Через открытые отверстия должны проваливаться кубики или шарики, попадая в желоба, а там, где они закрыты, ничего происходить не должно. Его [массив отверстий] можно сдвигать из одной колонки в другую по мере необходимости».

напоминала бильярдный стол с шарами, лузами, киями и желобами, и она открывала лузы в зависимости от последовательности единиц («открыто») и нулей («закрыто»).

«Пошаговая машина выводов» Лейбница заложила основу для новых теорий, включая ту идею, что, если логическое мышление может быть сведено к символам и результат проанализирован как вычислительная система, если геометрические задачи могут быть решены при помощи символов и чисел, тогда все может быть сведено к битам, включая человеческое поведение. Это значительный отход от идей философов прошлого: машины будущего станут воспроизводить человеческое мышление, не нуждаясь в идее божественного провидения. Для мышления не обязательно необходимы восприятие, чувства или душа. Лейбниц представлял себе компьютер, способный решать задачи общего характера, даже не обязательно математические. Он предполагал, что язык может быть сведен к элементарным идеям математики и науки и эту задачу можно рассматривать как часть работы над универсальным переводчиком²².

Может быть, ум и машина просто следуют алгоритму?

Если Лейбниц был прав в том, что люди – просто одушевленные машины, и когда-нибудь будут изобретены машины неодушевленные, способные производить непредставимые нам и отточенные идеи, тогда в нашем мире должно существовать простое и однозначное разделение машин на два класса: мы и они. Но спор только начался.

В 1738 году Жак де Вокансон, художник и изобретатель, построил несколько автоматов для Французской академии наук, и среди них – сложно устроенную утку, очень похожую на живую. Она не только подражала движениям настоящей утки, хлопая крыльями и клюя зерно, но и умела имитировать пищеварение. Таким образом, философам была предложена пища для размышлений: если нечто выглядит как утка и крикает как утка, является ли оно на самом деле уткой? Если мы полагаем, что утка обладает душой, отличной от нашей, каково достаточное условие для вывода, что утка осознает себя, и что из этого следует?

Шотландский философ Дэвид Юм отвергал идею, что осознание своего существования служит доказательством наличия сознания. В отличие от Декарта, Юм был сторонником эмпиризма. Он разработал научный метод, основанный на наблюдаемых фактах и логических рассуждениях. Когда Вокансон показывал свою оснащенную пищеварительной системой утку – и задолго до того, как впервые заговорили об ИИ, – Юм писал в «Трактате о человеческой природе»: «Разум является, и должен быть, всего лишь слугой страстей». В данном случае под «страстями» Юм подразумевал «иррациональные мотивы» и утверждал, что наше поведение определяют привлекательные для нас цели, а не абстрактная логика. Если наши впечатления – просто восприятие нами вещей, которые мы видим, чувствуем, осязаем, пробуем на вкус, обоняем, а идеи – восприятие вещей, с которыми мы не входим в непосредственное соприкосновение, то наше существование и картина окружающего нас мира, по Юму, оказываются основаны на концепции человеческого восприятия.

Разработка сложных автоматов, облик которых приобретал все большую реалистичность, более внимательное рассмотрение идеи компьютеров как мыслящих машин привели французского врача и философа Жюльена Офре де Ламетри к мысли предпринять исследование людей, животных и автоматов – столь же радикальное, сколь и скандальное. В статье 1747 года, которую он сначала опубликовал анонимно, Ламетри утверждал, что люди удивительно похожи на

²² Лейбниц пишет: «Я еще раз задумался над своим давнишним планом создания системы логических выводов, которая могла бы служить средством общения для всех разнообразных народов... Если бы мы располагали подобным универсальным устройством, то могли бы обсуждать проблемы метафизики или этики так же, как проблемы и задачи по математике или геометрии. В этом и заключалась моя цель: любое непонимание стало бы не более чем ошибкой в вычислениях... легко устранимой при помощи грамматических правил нового языка. Таким образом, в случае трудного спора философы могли бы просто сесть за стол и произвести вычисления. Подобно двум математикам, они могли бы сказать: „Давайте это подсчитаем“».

животных и обезьяна могла бы освоить человеческий язык, если бы ее «надлежащим образом учили». Ламетри также сделал вывод, что люди и животные – просто машины, управляемые инстинктом и опытом. «Человеческое тело – машина, которая сама заводит свои пружины; душа – не более чем принцип движения или материальная и осязаемая часть мозга»²³.

Из мысли, что люди представляют собой просто машины, приводимые в действие материальными силами, – то есть шестерни и колеса, выполняющие определенный набор функций, – следовало, что мы не особенные и не уникальные. Из нее также следовало, что, может быть, нас можно программировать. Окажись это правдой, от конструирования убедительно выглядящих уток и миниатюрных монахов люди когда-нибудь смогут перейти к изготовлению копий самих себя – и делать разнообразные разумные, мыслящие машины.

Можно ли построить мыслящую машину?

К 1830-м годам математики, инженеры и ученые всерьез взялись за постройку машины, способной производить вычисления так же, как люди-«компьютеры». Английский математик Ада Лавлейс и ученый Чарльз Бэббидж изобрели «разностную машину», а позже разработали проект более сложной «аналитической машины», решавшей математические задачи путем выполнения заранее определенной последовательности шагов. Бэббидж не предполагал, что его машина будет использована для чего-либо, кроме действий над числами. Именно Лавлейс в примечаниях к научной статье, которую тогда переводила, добавила изумительно глубокий комментарий, что более мощный вариант машины можно было бы использовать иначе²⁴. Если машина способна манипулировать символами, обозначающими разные вещи (например, музыкальные ноты), тогда ее можно было бы использовать для «размышления» о вещах, лежащих за пределами математики. Хотя Лавлейс не верила, что компьютер когда-либо обретет способность к самостоятельному мышлению, она предвидела создание сложной системы, умеющей выполнять инструкции и таким образом подражать человеку во многих его повседневных задачах.

В ста милях к северу от Кембриджского университета, где работали Лавлейс и Бэббидж, молодой математик-самоучка по имени Джордж Буль шел через поле в Донкастере, когда его внезапно осенило: он решил посвятить свою жизнь расширению логики человеческого мышления²⁵. Во время этой прогулки родилось то, что мы сегодня называем булевой алгеброй: способ упрощения логических выражений (например, «и», «или», «не») через использование символов и чисел. Скажем, вычисление выражения «истина и истина» должно давать результат «истина», что физически могло бы соответствовать положению переключателей или крышек луз на компьютере. Булю потребовалось два десятилетия на формализацию своих идей. И еще только через сто лет кому-то пришлось в голову, что булева логика в сочетании с теорией вероятностей могла бы превратить компьютеры из средства автоматизации элементарных математических операций в более сложные мыслящие машины. Технологии, позволяющей построить такую машину, еще не было – отсутствовали необходимые процессы, материалы и источники энергии, – и проверить теорию на практике было невозможно.

Переход от теоретического представления о мыслящей машине к компьютерам, начавшим имитировать мышление человека, произошел одновременно, с публикацией двух основополагающих статей: «Символический анализ релейных и переключательных схем» Клода Шеннона и «О вычислимых числах и их применении к проблеме разрешения» (*Entscheidungsproblem*) Алана Тьюринга. Изучая электротехнику в Массачусетском тех-

²³ “Apes to Androids: Is Man a Machine as La Mettrie Suggests?” http://www.charliemccarron.com/man_a_machine/.

²⁴ Luigi Manabrea, Sketch of the Analytical Engine Invented by Charles Babbage (London: Richard and John E. Taylor, 1843).

²⁵ Desmond MacHale, The Life and Work of George Boole: A Prelude to the Digital Age, New ed. (Cork University Press, 2014).

нологическом институте, Шеннон в качестве предмета по выбору взял философию, что выглядело необычно. Основным научным трудом, на который опиралась его диссертация, был трактат Джорджа Буля «Исследование законов мышления». Научный руководитель Шеннона, Вэнивар Буш, подал ему идею реализовать булеву логику в виде физических схем. Буш построил усовершенствованную версию «аналитической машины» Лавлейс и Бэббиджа – она называлась «дифференциальный анализатор», – но сконструирована она была в некоторой степени бессистемно. В то время не существовало теории, которая диктовала бы методику проектирования электрических схем. Открытие Шеннона заключалось в том, что он осуществил схемную реализацию булевой логики и объяснил, каким образом с ее помощью получить рабочую схему, способную складывать нули и единицы.

В то же время, когда Шеннон работал над переносом булевой логики на физические схемы, Тьюринг экспериментировал с «универсальным переводчиком» Лейбница, способным представлять все физическое и научное знание. Английский ученый ставил перед собой задачу доказать так называемую *Entscheidungsproblem*, или «проблему²⁶ разрешения». Упрощая, можно сформулировать ее так: не существует алгоритма, при помощи которого возможно доказать истинность или ложность произвольного математического утверждения. Ответ оказался отрицательным. Тьюринг сумел продемонстрировать, что такого алгоритма действительно не существует, но побочным результатом его работы явилась математическая модель универсальной вычислительной машины²⁷.

И это изменило все. Тьюринг понял, что программа и данные могут храниться внутри компьютера – для 1930-х годов это было радикальной идеей. До того все сходились на мысли, что машина, программа и данные – три независимые друг от друга сущности. Универсальная машина Тьюринга объясняла, почему они крепко связаны друг с другом. Если смотреть на вещи механически, логика управления схемами и переключателями тоже может быть закодирована в программе и данных. Подумайте на секунду о важности этих утверждений. Контейнер, программа и данные оказались объединены в рамках общей сущности подобно тому, как обстоит дело у людей. Мы тоже контейнеры (наши тела), программы (автономные функции клеток) и данные (наша ДНК в сочетании с прямой и косвенной информацией, поставляемой органами чувств).

В то же время давняя традиция конструирования автоматов, что началась с маленького монаха, умевшего ходить и молиться, наконец встретилась с работой Тьюринга и Шеннона. Американская промышленная компания Westinghouse построила робота, выполненного на базе электрических реле, для Всемирной выставки 1939 года. Робота звали *Elektro the Moto-Man*, «механический человек Электро». Он выглядел как гигантская человекоподобная фигура золотистого цвета и довольно грубой формы. Под ногами у него были колеса. Внутри находились 48 реле, используемых в телефонной коммутации.

Электро умел реагировать на команды, подаваемые ему голосом по телефонной трубке, воспроизводя заранее записанные ответы на встроенном проигрывателе. Он представлял собой антропоморфный компьютер, способный принимать элементарные решения – например, что сказать – без непосредственного участия человека в данный момент.

²⁶ В русскоязычной литературе устоялось название «проблема разрешения», но по существу эта задача представляет собой теорему, поэтому мы позволяем себе использовать в отношении ее глагол «доказать». – *Прим. пер.*

²⁷ Лучшее объяснение содержится в книге ученого-логика Мартина Дэвиса *The Universal Computer: The Road from Leibniz to Turing*: «Тьюринг знал, что лучше всего задавать алгоритм в виде списка правил, которые человек мог бы исполнять механически, подобно рецепту в кулинарной книге. Он сумел показать, что список действий для такого человека может быть сведен к небольшому количеству исключительно простых операций без последствий для результата вычисления. Затем, доказав, что способная выполнять только эти элементарные операции машина могла бы определить, следует ли данное заключение из указанных предпосылок... он пришел к выводу, что алгоритма, удовлетворяющего условиям *Entscheidungsproblem*, не существует».

Из газетных заголовков, научно-фантастических рассказов и кинохроники того времени видно, что люди были застигнуты врасплох, потрясены и напуганы подобным развитием событий. Для них происходящее выглядело так, словно эра «мыслящих машин» во всей своей полноте наступила в один день. Фантаст Айзек Азимов опубликовал пророческий рассказ «Лжец» в выпуске журнала *Astounding Science Fiction* за май 1941 года. Рассказ был его реакцией на достижения науки, свидетелем которых он оказался, и в нем Азимов впервые сформулировал знаменитые Три закона робототехники²⁸.

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

2. Робот должен повиноваться командам человека, если эти команды не противоречат Первому закону. 3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму законам.

Позже Азимов добавил так называемый Нулевой закон.

0. Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинен вред.

Мыслит ли мыслящая машина?

В 1943 году ученые-психиатры Уоррен Маккалоу и Уолтер Питтс из Чикагского университета опубликовали важную статью «Логическое исчисление идей, присущих нервной деятельности», где описывалась система нового типа, моделирующая живые нейроны при помощи нейронной сети простой архитектуры. Если контейнеры, программы и данные крепко связаны между собой, как утверждал Тьюринг, и если люди представляют собой аналогичные элегантно сконструированные контейнеры, следовательно, можно построить мыслящую машину, если смоделировать часть человеческого тела, ответственную за мышление, – мозг. Они сформулировали современную вычислительную теорию разума и мозга, «нейронную сеть». Вместо того чтобы сосредоточиться на идеях аппаратного и программного обеспечения, они предложили систему нового типа, способную перерабатывать огромные объемы данных, в точности как это делаем мы. Мощности компьютеров еще не хватало, чтобы проверить их теорию, но статья вдохновила других, начавших работу над разумными компьютерными системами.

Связь между разумными компьютерными системами и автономным принятием решений прояснилась благодаря публикации обширного трактата по прикладной математике, вышедшего из-под пера Джона фон Неймана, американского полимата венгерского происхождения, специалиста в области кибернетики, физики и математики. В 1944 году Нейман в соавторстве с Оскаром Моргенштерном, экономистом из Принстонского университета, выпустил книгу объемом в 640 страниц, где подробнейшим образом рассказывалось, как теория игр объясняет основы любых экономических решений²⁹. Именно благодаря этой работе фон Неймана пригласили сотрудничать с Вооруженными силами США, в то время занимавшимися разработкой электрической вычислительной машины нового типа, «Электронного числового интегратора и вычислителя», сокращенно ЭНИАК. Вначале инструкции, предназначенные для машины, реализовывались постоянными соединениями внутри нее, поэтому с каждой новой программой соединения во всей системе нужно было переделывать заново. Вдохновленный трудами Тьюринга, Маккалоу и Питтса, фон Нейман разработал новую систему хранения программ на самом компьютере. Тем самым совершился переход от первой эпохи развития вычислительной техники (табуляции) к новой – эпохе программируемых систем.

²⁸ На самом деле в рассказе «Лжец» сформулирован только Первый закон. Все три впервые появляются в рассказе «Хоро-вод». – Прим. пер.

²⁹ Русское издание: Дж. фон Нейман, О. Моргенштерн «Теория игр и экономическое поведение». М., Наука, 1970. – Прим. пер.

Сам Тьюринг к этому моменту работал над концепцией нейронной сети, составленной из компьютеров, каждый из которых был выполнен по архитектуре с хранимым программным кодом. *London Times* цитировала Тьюринга в 1949 году: «Я не вижу причин, почему она (машина) не могла бы войти в какую-либо сферу профессиональной деятельности, обычно занятую людьми, и постепенно начать конкурировать с ними на равных. Не думаю, что можно исключить даже написание сонетов, хотя здесь сравнение становится немного нечестным: ведь сонет, написанный машиной, скорее оценит по достоинству другая машина». Годом позже, в статье для философского журнала *Mind*, Тьюринг попытался ответить на вопросы, поставленные Гоббсом, Декартом, Юмом и Лейбницем. В ней он предложил тезис и тест: если когда-нибудь компьютер окажется способен отвечать на вопросы неотличимым от человека образом, он должен быть «мыслящим». Вы наверняка слышали об этой статье под другим названием: «Тест Тьюринга».

Статья начиналась со ставшего знаменитым вопроса, который с тех пор многократно задавали – и на который отвечали – столь многие философы, теологи, математики и ученые до него: «Могут ли машины мыслить?» Но Тьюринг, помнивший о многовековых спорах об уме и машине, отбрасывает этот вопрос как чересчур общий для осмысленной дискуссии. «Машина» и «мыслить» – слишком общие слова, оставляющие слишком много места для субъективной интерпретации (в конце концов, 400 лет люди писали книги и статьи, посвященные их значению).

Игра, по Тьюрингу, строилась на обмане, и компьютер «выигрывал», когда его оказывалось невозможно отличить от человека. Тест проводится следующим образом: имеются человек, машина и находящееся в отдельной комнате лицо, задающее вопросы. Цель игры для лица, задающего вопросы, заключается в том, чтобы понять, какие ответы даются человеком, а какие – машиной. Перед началом игры лицо, задающее вопросы, получает обозначения для своих собеседников, *X* и *Y*, но не знает, какая из них относится к компьютеру, и может только задавать вопросы типа: «Не мог бы *X* сообщить мне, играет ли он в шахматы?» В конце игры ему необходимо указать, кем был *X* и кем *Y*. Задачей второго человека является помочь лицу, задающему вопросы, определить машину, а задачей машины – обмануть задающего вопросы так, чтобы тот поверил, что она, машина, и есть второй человек. О самой игре Тьюринг писал: «Я полагаю, что приблизительно через пятьдесят лет станет возможным запрограммировать компьютеры с запоминающим устройством емкостью примерно³⁰ 10^9 играть в имитационную игру так, что для среднего лица, задающего вопросы, шанс идентифицировать стороны правильно после пяти минут собеседования не превысит 70 %»³¹.

Но Тьюринг был ученым и отдавал себе отчет в том, что его теорию невозможно доказать, – по крайней мере при его жизни. Как оказалось, проблема заключалась не в отсутствии эмпирических свидетельств, показывающих, что машины когда-нибудь обретут возможность мыслить, и даже не во времени жизни ученого – сам Тьюринг говорил, что провести его тест, скорее всего, окажется возможно только к концу двадцатого столетия. «Мы можем надеяться, что машины со временем станут конкурировать с людьми во всех чисто интеллектуальных сферах деятельности», – писал Тьюринг. Настоящая проблема заключалась в том, чтобы совершить рискованный шаг, поверив, что когда-нибудь машины смогут видеть, рассуждать и запоминать, и люди могут оказаться помехой на пути их прогресса. Его коллегам-исследователям тогда пришлось бы наблюдать способность к познанию, лишённую духовной природы, поверить в возможность существования разумных машин, которые, в отличие от людей, станут принимать решения бессознательно.

³⁰ В тексте отсутствует указание, в каких единицах выражается емкость запоминающего устройства. – *Прим. пер.*

³¹ Alan Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433–60.

Лето и зима ИИ

В 1955 году профессора Марвин Минский (математика и неврология), Джон Маккарти (математика), а также Клод Шеннон (математик и криптограф из Bell Labs) и Натаниэль Рочестер (специалист по информатике из IBM) предложили провести двухмесячный семинар, посвященный работам Тьюринга и перспективам машинного обучения. Их идея заключалась в следующем: если возможно описать каждое проявление умственной деятельности человека, то можно и научить машину имитировать ее³². Но для этого потребовалась бы многочисленная группа экспертов в разных областях человеческой деятельности. Они полагали, что, собрав междисциплинарную группу исследователей и интенсивно работая без перерывов в течение целого лета, можно добиться существенного прогресса.

Состав группы имел решающее значение. В нее вошли первоклассные инженеры, социологи, психологи, математики, физики и специалисты по когнитивистике. Им предстояло ставить фундаментальные вопросы и отвечать на них: что значит «мыслить», как работает наш «разум» и как научить машины учиться таким же образом, каким это делают люди? Идея заключалась в том, что разноплановая группа продолжит совместную работу после завершения семинара, развивая эту новую область знания. Поскольку речь шла о междисциплинарном подходе к созданию мыслящих машин, для его обозначения понадобилось новое имя. Участники остановились на немного двусмысленном, но элегантном термине «искусственный интеллект».

Маккарти составил предварительный список из 47 экспертов, присутствие которых считал необходимым, чтобы завязать нужные связи между людьми и заложить основу для дальнейших исследований и работ по прототипированию. Составить список людей, без которых нельзя было обойтись, когда по-настоящему начнутся работы по концептуализации и конструированию ИИ, было очень сложно. В частности, Минскому не нравилось, что на встрече не будет двух ключевых фигур: Тьюринг умер за два года до того, а фон Нейман находился в терминальной стадии рака³³.

³² “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” Stanford Computer Science Department’s Formal Reasoning Group, John McCarthy’s home page, links to articles of historical interest, last modified April 3, 1996, <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.

³³ В своем предложении Маккарти, Минский, Рочестер и Шеннон привели такой список экспертов, которых намеревались пригласить в Дартмутский университет для работы над ИИ. Я воспроизвожу список так, как он был опубликован в 1955 году, с указанием компаний и адресов. Не все упомянутые в нем лица смогли принять приглашение. А-дельсон, Марвин Hughes Aircraft Company Эрпорт-Стейшен, Лос-Анджелес, Калифорния Эшби, В. Р. Барнвуд-хаус Глостер, Англия Бэкус, Джон IBM Corporation Мэдисон-авеню, 590 Нью-Йорк, штат Нью-Йорк Бернстайн, Алекс IBM Corporation Мэдисон-авеню, 590 Нью-Йорк, штат Нью-Йорк Бигелу, Дж. Х. Институт перспективных исследований Принстон, Нью-Джерси Элиас, Питер Научно-исследовательская лаборатория электроники Массачусетский технологический институт (МТИ) Кембридж, Массачусетс Дьюда, В. Л. Научно-исследовательская лаборатория IBM Покипси, штат Нью-Йорк Дэвис, Пол М. 18-я улица, 1317 С. Лос-Анджелес, Калифорния Фэно Р. М. Научно-исследовательская лаборатория электроники МТИ Кембридж, Массачусетс Фэрли, Б. Г. Парк-авеню, 324 Арлингтон, Массачусетс Галантер, Э. Х. Пенсильванский университет Филадельфия, штат Пенсильвания Гелернтер, Герберт Научно-исследовательская лаборатория IBM Покипси, штат Нью-Йорк Гласхоу, Харви А. Оливия-стрит, 1102 Энн-Арбор, Мичиган Гёрцаль, Герберт Западная 11-я улица, 330 Нью-Йорк, штат Нью-Йорк Хагельбаргер, Д. Научно-исследовательская лаборатория Bell Telephone Мюррей-Хилл, Нью-Джерси Миллер Джордж А. Мемориал Холл Гарвардский университет Кембридж, Массачусетс Хармон, Леон Д. Научно-исследовательская лаборатория Bell Telephone Мюррей-Хилл, Нью-Джерси Холланд, Джон Х. Научно-исследовательский институт окружающей среды Мичиганский университет Энн-Арбор, Мичиган Холт, Анатолий Рурал-лейн, 7358 Филадельфия, штат Филадельфия Каутц, Уильям Х. Стэнфордский научно-исследовательский институт Менло-Парк, Калифорния Люс Р. Д. Западная 117-я улица, 427 Нью-Йорк, штат Нью-Йорк Маккей, Дональд Физический факультет Лондонский университет Лондон, почтовый индекс «Западный Центральный 2 (WC2)», Англия Маккарти, Джон Дартмутский университет Хановер, Нью-Гэмпшир Маккулло, Уоррен С. Научно-исследовательская лаборатория электроники, МТИ Кембридж, Массачусетс Мелзак З. Э. Математический факультет Мичиганский университет Энн-Арбор, Мичиган Минский М. Л. Ньюбери-стрит, 112 Бостон, Массачусетс Мор, Тренчард Факультет информационных технологий МТИ Кембридж, Массачусетс Нэш, Джон Институт перспективных исследований Принстон, Нью-Джерси Ньюэлл, Аллен Факультет промышленного управления Технологический институт Карнеги Питтсбург, Пенсильвания Робинсон, Абрахам Математический факультет Торонтский университет Торонто, Онтарио, Канада Рочестер,

Тем не менее в своих стараниях собрать разностороннюю группу специалистов, чьи знания и навыки взаимно дополняли бы друг друга, организаторы упустили из виду несколько существенных аспектов. Все в списке были белыми – несмотря на то, что в областях знания, интересовавших Маккарти и Минского, работало множество первоклассных цветных специалистов. Все они работали или в технологических гигантах своего времени (IBM, Bell Labs), или в узком кругу избранных университетов. Множество одаренных женщин уже тогда вносили заметный вклад в инженерное дело, информатику, математику, физику – но им места в списке не нашлось³⁴. Среди приглашенных были только мужчины, за исключением жены Марвина Минского, Глории. Не отдавая себе отчета в своих предубеждениях, эти ученые, пытавшиеся понять, как устроен человеческий ум, как мы мыслим и как машины могли бы учиться у всего человечества, резко ограничили свою выборку теми, кто выглядел и говорил так же, как они.

В следующем году группа собралась на последнем этаже факультета математики Дартмутского колледжа. Среди тем ее работ были теория сложности, моделирование естественного языка, нейронные сети, связь случайности и творчества, а также самообучающиеся машины. По рабочим дням они собирались в главной аудитории факультета для общего обсуждения, после которого расходились, чтобы работать над частными задачами. Во время одного из таких общих собраний профессора Алан Ньюэлл, Герберт Саймон и Клифф Шоу предложили способ доказательства теорем логики и смоделировали процесс вручную. Свою программу они назвали Logic Theorist, «Ученый логик». Она оказалась первой программой, способной имитировать навыки решения задач, присущие человеку. (Впоследствии программа доказала 38 из 52 теорем из «Оснований математики» Альфреда Норта Уайтхеда и Бертрана Рассела,

Натаниэль Научно-техническая исследовательская лаборатория IBM Corporation Покипси, штат Нью-Йорк Роджерс, Хартли, мл. Математический факультет МТИ Кембридж, Массачусетс Розенблит, Волтер Научно-исследовательская лаборатория электроники МТИ Кембридж, Массачусетс Ротштейн, Джером Ист-Берген-Плейс, 21 Ред Бэнк, Нью-Джерси Соир, Дэвид IBM Corporation Мэдисон-авеню, 590 Нью-Йорк, штат Нью-Йорк Шорр-Кон, Дж. Дж. С-380 Lincoln Laboratory, MIT Lexington, МАШЭпли, J. Rand Corporation Мэйн-стрит, 1700 Санта-Моника, Калифорния Шутценбергер, М. П. Научно-исследовательская лаборатория электроники МТИ Кембридж, Массачусетс Селфридж, О. Дж. Лаборатория Линкольна МТИ Лексингтон, Массачусетс Шэннон, К. Е. Научно-исследовательская лаборатория электроники МТИ Кембридж, Массачусетс Шапиро, Норман Rand Corporation Мэйн-стрит, 1700 Санта-Моника, Калифорния Саймон, Герберт А. Факультет промышленного управления Технологический институт Карнеги Питтсбург, Пенсильвания Соломонофф, Раймонд Дж. Technical Research Group Юнион-сквер запад, 17 Нью-Йорк, штат Нью-Йорк Стил Дж. Э. капитан ВВС США Зона Б, почтовый ящик 8698 База ВВС Райт-Паттерсон Огайо Уэбстер, Фредерик Кулидж-авеню, 62 Кембридж, Массачусетс Мур Э. Ф. Научно-исследовательская лаборатория Bell Telephone Мюррей-Хилл, Нью-Джерси Кемени Джон Дж. Дартмутский университет ХанOVER, Нью-Гэмпшир

³⁴ Я составила очень краткий список женщин и цветных, которые могли бы внести большой вклад в работу семинара в Дартмутском университете, но были обойдены вниманием организаторов. Этот список ни в коем случае не исчерпывающий. Я могла бы продолжать в течение многих десятков страниц. Перед вами же – просто выборка способных, талантливых, творческих людей, оставшихся за границами списка. Джеймс Эндрюс, математик, профессор Университета штата Флорида, специализировавшийся на теории групп и теории узлов. • Джин Бартик, математик, одна из первых программистов компьютера ЭНИАК. Альберт Тернер Баруха-Рейд, математик и теоретик, внесший существенный вклад в теорию марковских цепей, теорию вероятностей и статистику. • Дэвид Блэкуэлл, статистик и математик, сделавший существенный вклад в теорию игр, теорию информации, теорию вероятностей и байесову статистику. • Мэми Фиппс Кларк, PhD и социальный психолог, чья исследовательская работа была посвящена самосознанию. • Тельма Эстрин, пионер использования компьютерных систем в нейрофизиологии и науке о мозге. Она была научным сотрудником департамента электроэнцефалографии Института неврологии при Пресвитерианской больнице в Нью-Йорке. • Эвелин Бойд Грэнвилл, PhD по математике, разработчик компьютерных программ анализа траектории, предназначенных для первых пилотируемых орбитальных и лунных космических полетов США. • Бетти Холбертон, математик и одна из первых программистов компьютера ЭНИАК. Изобрела точки останова, применяемые в отладке компьютерных программ. • Грейс Хоппер, специалист по информатике и создатель языка Кобол, одного из первых языков программирования, применяемого и по сей день. • Мэри Джексон, инженер и математик, позже ставшая первой черной женщиной-инженером в НАСА. • Кэтлин Макналти, математик и одна из первых программистов компьютера ЭНИАК. • Мэрлин Мельцер, математик и одна из первых программистов компьютера ЭНИАК, первой программируемой вычислительной машины, выполненной полностью на электронной элементной базе. • Рожа Петер, математик и основатель теории рекурсивных функций. • Фрэнсис Спенс, математик и одна из первых программистов компьютера ЭНИАК. • Рут Тейтельбаум, математик и одна из первых программистов компьютера ЭНИАК. Совместно с Мэрлин Мельцер разработала уравнения для расчета баллистической траектории. • Дороти Вон, математик и человек-«компьютер», в 1949 году исполнявшая обязанности директора West Area Computers. • Джесси Эрнст Уилкинс-мл., специалист по ядерной физике, инженер-механик и математик, ставшая самым молодым студентом Чикагского университета в возрасте 13 лет.

фундаментального учебника по началам математической науки.) Клод Шеннон, несколькими годами ранее учивший компьютеры играть в шахматы против людей, получил возможность показать прототип своей программы, работа над которой тогда еще продолжалась³⁵.

Тем летом в Дартмуте расчетам Маккарти и Минского на фундаментальный прорыв в области ИИ не суждено было оправдаться. Чтобы превратить ИИ из теории в практику, им не хватило времени, не говоря уже о вычислительной мощностии³⁶. Тем не менее именно к этому моменту восходят три практики, образующие фундамент ИИ в известном нам сегодня виде:

– теоретическая и практическая разработка ИИ, его тестирование и совершенствование будут производиться крупными технологическими компаниями и академической наукой, работающими совместно; – совершенствование ИИ требует больших сумм денег, поэтому станет необходима коммерциализация этих работ в той или иной форме – будь то взаимодействие с государственными или военными органами через партнерства или разработка продуктов и систем, которые могут быть проданы;

– исследование и разработка ИИ будут опираться на сеть ученых, работающих на стыке дисциплин, что означает создание новой отрасли науки с нуля. Это также означает, что уже работающие в этой отрасли будут нанимать, как правило, знакомых им людей, обеспечивая относительную однородность сети и ограничивая выборку.

Тем летом произошло другое интересное событие. Работая над вопросом Тьюринга, могут ли машины мыслить, группа разделилась из-за подходов к ответу на него, а именно построению самообучающейся машины. Некоторые участники предпочитали биологический метод. Иными словами, они считали, что при помощи нейронных сетей можно придать ИИ здравый смысл и способность рассуждать логически – и таким образом машины могут стать разумными. Другие возражали, что построить настолько полную копию мыслительного аппарата человека невозможно никогда. Вместо этого они предлагали инженерный подход. Вместо того чтобы подавать машине команды, приводящие к решению задачи, программа могла бы помочь системе «обучиться» на наборе данных. Система делала бы предсказания, основываясь на этих данных, а контролирующей ее человек проверял бы ответы, тренируя и настраивая ее в процессе работы. Таким образом определяется «машинное обучение» в узком смысле, то есть обучение решению изолированной задачи, например игре в шашки.

Психолог Фрэнк Розенблатт, участвовавший в семинаре в Дартмутском колледже, хотел смоделировать обработку визуальной информации в человеческом мозгу и в итоге научить машину распознавать предметы. Он рассчитывал создать простую рамочную программу, восприимчивую к обратной связи. Опираясь на результаты летних исследований, Розенблатт создал систему, которую назвал Perceptron. Это была первая искусственная нейронная сеть (ANN), работавшая по принципу создания связей между многочисленными элементами, обрабатывающими информацию и расположенными послойно. Каждый механический нейрон принимал на входе множество различных сигналов и затем обрабатывал их математически с учетом весов, чтобы определить, какой сигнал генерировать на выходе. Такая параллельная структура позволяла осуществлять доступ ко множеству обрабатываемых элементов одновременно, а это означало, что она не только была быстрой, но и могла непрерывно обрабатывать большой объем данных.

³⁵ “The Dartmouth Workshop – as Planned and as It Happened,” Stanford Computer Science Department’s Formal Reasoning Group, John McCarthy’s home page, lecture “AI: Past and Future,” last modified October 30, 2006, <http://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html>.

³⁶ “The Dartmouth AI Archives,” RaySolomonoff.com, <http://raysolomonoff.com/dartmouth/>.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

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

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