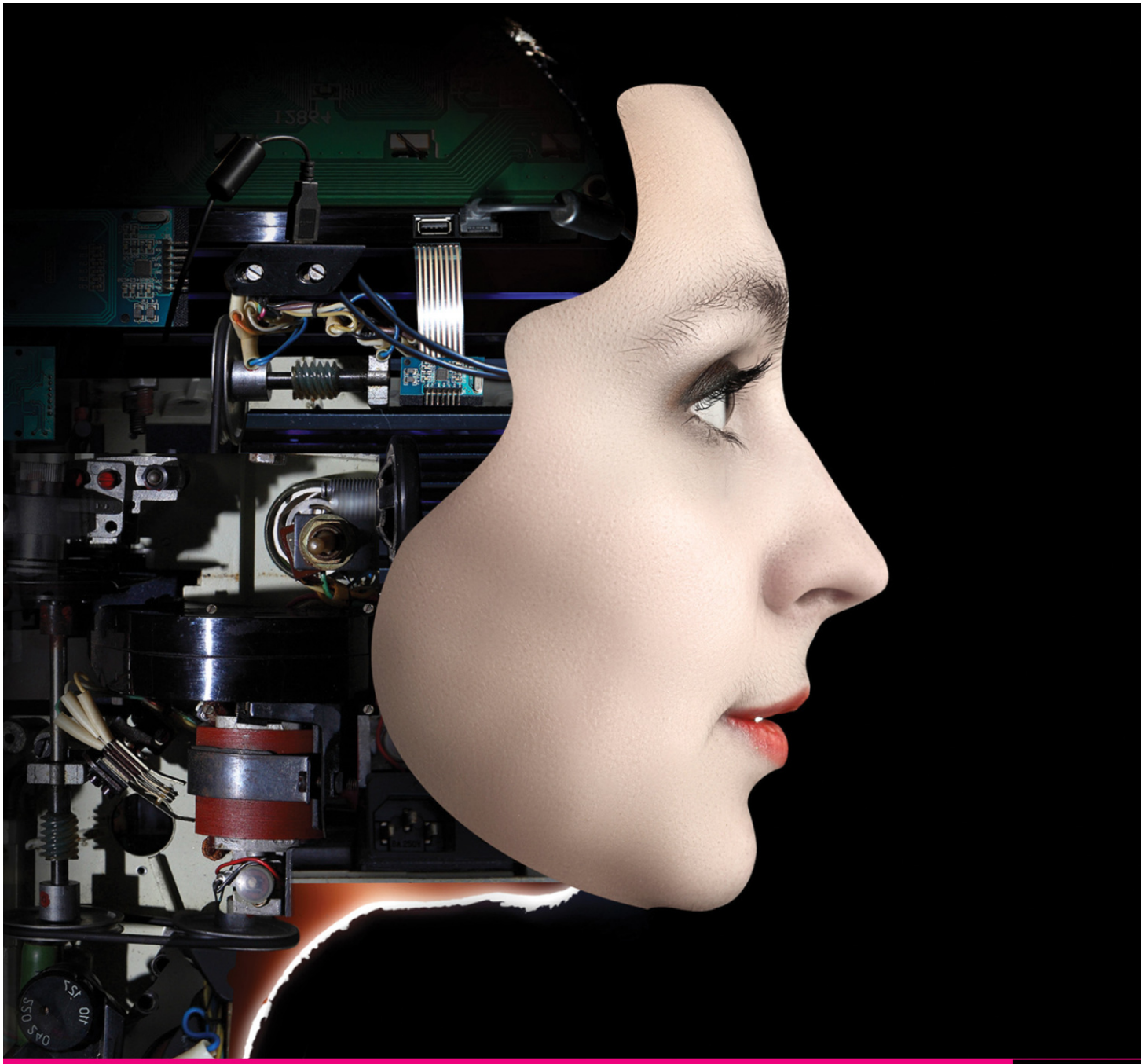

МЕЛАНИ
МИТЧЕЛЛ

ИДИОТ ИЛИ ГЕНИЙ?

КАК РАБОТАЕТ И НА ЧТО СПОСОБЕН ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Прежде чем машины заявят о собственной человечности...

Книжные проекты
Дмитрия Зимина

CORPUS

ЭЛЕМЕНТЫ 2.0

Элементы 2.0

Мелани Митчелл

Идиот или гений? Как работает и на что способен искусственный интеллект

«Corpus (АСТ)»
2019

УДК 004.8
ББК 32.813

Митчелл М.

Идиот или гений? Как работает и на что способен искусственный интеллект / М. Митчелл — «Corpus (АСТ)», 2019 — (Элементы 2.0)

ISBN 978-5-17-127256-2

За 65 лет, прошедших после Дартмутского семинара, который положил начало разработке искусственного интеллекта, в этой области совершено множество прорывов, однако до создания машины с «человеческим» интеллектом по-прежнему далеко. Сегодня ИИ распознает изображения и переводит речь, управляет беспилотными автомобилями, обыгрывает человека в шахматы и го, но пока не способен переносить навыки на новые задачи, может перепутать соль с дорожной разметкой, а автобус – со страусом. Мелани Митчелл, одна из ведущих ученых-информатиков, знакомит читателя с историей развития ИИ и принципами его работы, рассказывает о главных проблемах его применения и перспективах создания ИИ «человеческого уровня». В формате PDF A4 сохранён издательский дизайн.

УДК 004.8
ББК 32.813

ISBN 978-5-17-127256-2

Содержание

Пролог 6
ИИ и “ГЭБ” 7
Шахматы и первое зерно сомнения 9
Музыка – бастион человечности 11
Google и сингулярность 13
Почему Хофштадтер испуган? 14
Я в недоумении 15
О чем эта книга 17
Часть I 18
Глава 1 18
Два месяца и десять мужчин в Дартмуте 18
Основные понятия и работа с ними 19
Анархия методов 20
Символический ИИ 21
Субсимволический ИИ: перцептроны 22
Входные сигналы нашего перцептрона 24
Как узнать веса и порог перцептрона 25
Ограниченность перцептронов 27
Зима ИИ 29
Конец ознакомительного фрагмента. 30

Мелани Митчелл
Идиот или гений? Как работает и на что способен искусственный интеллект

Посвящается моим родителям, которые научили меня быть думающим человеком и много чему еще

MELANIE MITCHELL
ARTIFICIAL INTELLIGENCE
A GUIDE FOR THINKING HUMANS

Издательство CORPUS ®

Книжные проекты Дмитрия Зимина

Эта книга издана в рамках программы “Книжные проекты Дмитрия Зимина” и продолжает серию “Библиотека фонда «Династия»”. Дмитрий Борисович Зимин – основатель компании “Вымпелком” (*Beeline*), фонда некоммерческих программ “Династия” и фонда “Московское время”.

Программа “Книжные проекты Дмитрия Зимина” объединяет три проекта, хорошо знакомых читательской аудитории: издание научно-популярных книг “Библиотека фонда «Династия»”, издательское направление фонда “Московское время” и премию в области русскоязычной научно-популярной литературы “Просветитель”.

Подробную информацию о “Книжных проектах Дмитрия Зимина” вы найдете на сайте **ziminbookprojects.ru**

Пролог
Испуганные

Складывается впечатление, что компьютеры с пугающей скоростью становятся все умнее, но кое-что им все же не под силу – они по-прежнему не могут оценить иронию. Именно об этом я думала несколько лет назад, когда по пути на встречу по вопросам искусственного интеллекта (ИИ) заблудилась в столице поиска – в *Googleplex*, штаб-квартире *Google* в Маунтин-Вью, в штате Калифорния. Более того, я заблудилась в здании *Google Maps*. Это была ирония судьбы в квадрате.

Найти само здание *Google Maps* оказалось несложно. У двери стояла машина *Google Street View*, на крыше которой громоздилась установка с камерой, похожей на черно-красный футбольный мяч. Оказавшись внутри и получив у охраны хорошо заметный гостевой пропуск, я растерялась и заблудилась в лабиринтах кабинок, занятых стайками работников *Google*, которые, надев наушники, сосредоточенно стучали по клавишам компьютеров *Apple*. В конце концов мои бессистемные поиски (в отсутствие плана здания) увенчались успехом: я нашла нужную переговорную и присоединилась к собравшимся на целый день.

Эту встречу в мае 2014 года организовал Блез Агуэра-и-Аркас, молодой специалист по компьютерным технологиям, который недавно ушел с высокой должности в *Microsoft*, чтобы возглавить разработку машинного интеллекта в *Google*. В 1998 году компания *Google* начала с одного "продукта" – веб-сайта, на котором использовался новый, исключительно успешный метод поиска в интернете. С годами *Google* превратилась в главную технологическую компанию мира и теперь предлагает широкий спектр продуктов и услуг, включая *Gmail*, "*Google* Документы", "*Google* Переводчик", *YouTube, Android* и множество других, которыми вы пользуетесь каждый день, и таких, о которых вы, может, и не слышали вовсе.

Основатели *Google*, Ларри Пейдж и Сергей Брин, долгое время лелеяли мысль о создании искусственного компьютерного интеллекта, и эта сверхзадача предопределила одно из главных направлений деятельности *Google*. За последние десять лет компания наняла множество специалистов в области искусственного интеллекта, включая знаменитого изобретателя и противоречивого футуролога Рэя Курцвейла, который утверждает, что в ближайшем будущем наступит технологическая сингулярность, то есть момент, когда компьютеры станут умнее людей. Курцвейла наняли в *Google*, чтобы он помог воплотить эту идею в жизнь. В 2011 году внутри *Google* создали группу исследования ИИ, которую назвали *Google Brain*, а после этого компания приобрела впечатляющее количество ИИ-стартапов со столь же многообещающими названиями, среди которых *Applied Semantics, DeepMind* и *Vision Factory*.

Иными словами, *Google* уже нельзя назвать просто поисковым сервисом – даже с натяжкой. *Google* быстро превращается в компанию прикладного ИИ. Именно искусственный интеллект, подобно клею, связывает различные продукты, услуги и смелые исследования *Google* и его головной организации, *Alphabet*. Идеальная цель компании отражена в оригинальной концепции входящей в ее состав группы *DeepMind*: "Постичь интеллект и использовать его, чтобы постичь все остальное"[1].

[1] A. Cuthbertson, "DeepMind AlphaGo: AI Teaches Itself «Thousands of Years of Human Knowledge» Without Help", *Newsweek*, Oct. 18, 2017, *www.newsweek.com/deepmind-alphago-ai-teaches-human-help*-687620.

ИИ и “ГЭБ”

Устроенной в *Google* встречи по вопросам ИИ я ждала с нетерпением. Я начала работать с различными аспектами ИИ еще в аспирантуре, в 1980-х годах, а достижения *Google* меня восхищали. У меня также был целый ряд неплохих идей, которыми я хотела поделиться. Но я должна признать, что встречу собрали не ради меня. Ее организовали, чтобы лучшие исследователи ИИ из *Google* смогли побеседовать с Дугласом Хофштадтером, легендой ИИ и автором знаменитой книги с загадочным названием “Гёдель, Эшер, Бах: эта бесконечная гирлянда”, или коротко: “ГЭБ”. Если вы специалист в области компьютерных наук или просто увлекаетесь компьютерами, вероятно, вы слышали об этой книге, читали ее или пытались прочесть.

В написанной в 1970-х “ГЭБ” нашли отражение многие интеллектуальные увлечения Хофштадтера: обращаясь к математике, искусству, музыке, языку, юмору и игре слов, он поднимает глубокие вопросы о том, как разум, сознание и самосознание, неотъемлемо присущие каждому человеку, возникают в неразумной и бессознательной среде биологических клеток. В книге также рассматривается, как компьютеры в итоге смогут обрести разум и самосознание. Это уникальная книга – я не знаю ни одной другой, которая могла бы с ней сравниться. Вчитаться в нее нелегко, но она тем не менее стала бестселлером и принесла своему автору Пулитцеровскую премию и Национальную книжную премию. Несомненно, “ГЭБ” вдохновила заняться ИИ больше молодежи, чем любая другая книга. Среди этой молодежи была и я.

В начале 1980-х годов, окончив колледж, где я изучала математику, я обосновалась в Нью-Йорке, устроилась работать учителем и готовила подростков к поступлению в колледж. Я была несчастна и пыталась понять, чем мне на самом деле хочется заняться в жизни. Прочитав восторженный отзыв в журнале *Scientific American*, я узнала о “ГЭБ” и незамедлительно купила книгу. Следующие несколько недель я читала ее, все больше убеждаясь, что хочу не просто стать специалистом по ИИ, но и работать с Дугласом Хофштадтером. Никогда прежде я не приходила в такой восторг от книги и не чувствовала такой уверенности в выборе профессии.

В то время Хофштадтер преподавал информатику в Индианском университете. У меня родился отчаянный план подать документы на поступление в аспирантуру по информатике, приехать и убедить Хофштадтера принять меня на программу. Была лишь одна загвоздка: я никогда не изучала информатику. Я выросла в окружении компьютеров, потому что в 1960-х мой отец работал специалистом по ЭВМ в технологическом стартапе, а в свободное время собирал мейнфрейм в домашнем кабинете. На машине *Sigma* **2**, размером с холодильник, красовался значок “Я молюсь на фортране” – и я почти не сомневалась, что в ночи, когда все засыпают, компьютер и правда тихонько читает молитвы. Мое детство пришлось на 1960-е и 1970-е, а потому я постигла азы популярных в то время языков – сначала фортрана, а затем бейсика и паскаля, – но почти ничего не знала о методах программирования, не говоря уже о многих других вещах, которые должен знать человек, решивший поступить в аспирантуру по информатике.

Чтобы ускорить свое продвижение к цели, в конце учебного года я уволилась с работы, переехала в Бостон и стала изучать основы информатики, готовясь сменить карьеру. Через несколько месяцев после начала новой жизни я пришла на занятие в Массачусетский технологический институт (MIT) и заметила объявление о лекции Дугласа Хофштадтера, которая должна была два дня спустя состояться в том же кампусе. Я прослушала лекцию, дождалась своей очереди в толпе почитателей и сумела поговорить с Хофштадтером. Оказалось, что он приехал в MIT в разгар годичного академического отпуска, по окончании которого планировал перейти из Индианского университета в Мичиганский университет в Энн-Арбор.

Если не вдаваться в детали, я проявила настойчивость и убедила Хофштадтера взять меня на должность лаборанта – сначала на лето, а затем на оставшиеся шесть лет учебы в аспирантуре, после чего получила докторскую степень по информатике в Мичиганском университете. Мы с Хофштадтером продолжили тесно общаться, часто обсуждая ИИ. Он знал, что я интересуюсь исследованиями ИИ, которые проводятся в *Google*, и любезно предложил мне присоединиться к нему на встрече.

Шахматы и первое зерно сомнения

В переговорной, которую я так долго искала, собралось около двадцати человек (не считая нас с Дугласом Хофштадтером). Все они работали инженерами в *Google* и занимались исследованиями ИИ в составе разных команд. В начале встречи все по очереди представились. Несколько человек отметили, что занялись ИИ, потому что в юности прочитали “ГЭБ”. Им всем не терпелось услышать, что легендарный Хофштадтер скажет об ИИ. Затем Хофштадтер поднялся и взял слово. “Я хочу сказать несколько слов об исследованиях ИИ в целом и работе *Google* в частности, – начал он, преисполненный энтузиазма: – Я напуган. Очень напуган”.

Хофштадтер продолжил свое выступление[2]. Он рассказал, что в 1970-х, когда он приступил к работе над ИИ, все это казалось интересным, но таким далеким от жизни, что не было “ни опасности на горизонте, ни чувства, что все это происходит *на самом деле*”. Создание машин с человекоподобным интеллектом сулило чудесные интеллектуальные приключения в рамках долгосрочного исследовательского проекта, до воплощения которого, как говорили, оставалась по меньшей мере “сотня Нобелевских премий”[3]. Хофштадтер полагал, что теоретически ИИ возможен: “Нашими «врагами» были люди вроде Джона Сёрла, Хьюберта Дрейфуса и других скептиков, которые утверждали, что ИИ невозможен. Они не понимали, что мозг – это кусок вещества, которое подчиняется законам физики, а компьютер может моделировать что угодно… уровень нейронов, нейротрансмиттеров и так далее. Все это возможно в теории”. Идеи Хофштадтера о моделировании интеллекта на разных уровнях – от нейронов до сознания – подробно разбирались в “ГЭБ” и десятилетиями лежали в основе его собственных исследований. Но до недавних пор Хофштадтеру казалось, что на практике общий ИИ “человеческого уровня” не успеет появиться при его жизни (и даже при жизни его детей), поэтому он не слишком об этом беспокоился.

В конце “ГЭБ” Хофштадтер перечислил “Десять вопросов и возможных ответов” об искусственном интеллекте. Вот один из них: “Будут ли такие шахматные программы, которые смогут выигрывать у кого угодно?” Хофштадтер предположил, что таких программ не будет. “Могут быть созданы программы, которые смогут обыгрывать кого угодно, но они не будут исключительно шахматными программами. Они будут программами *общего* разума”[4].

На встрече в *Google* в 2014 году Хофштадтер признал, что был “в корне неправ”. Стремительное совершенствование шахматных программ в 1980-х и 1990-х годах посеяло первое зерно сомнения в его представления о краткосрочных перспективах ИИ. Хотя в 1957 году пионер ИИ Герберт Саймон предсказал, что компьютерная программа станет чемпионом мира “не позднее, чем через десять лет”, в середине 1970-х, когда Хофштадтер писал “ГЭБ”, лучшие компьютерные шахматные программы играли лишь на уровне хорошего (но не блестящего) любителя. Хофштадтер подружился с шахматным энтузиастом и профессором психологии Элиотом Хёрстом, который много писал об отличиях выдающихся шахматистов от шахматных программ. Эксперименты показали, что шахматисты обычно выбирают ход, быстро узнавая комбинацию на доске, а не перебирая все возможные варианты методом “грубой силы”, как делают компьютерные программы. В ходе партии лучшие шахматисты считают комбинацию фигур определенным “положением”, которое требует определенной “стратегии”. Иными словами, шахматисты быстро распознают конкретные комбинации и стратегии как элементы концепций более высокого уровня. Хёрст утверждал, что в отсутствие такой способности узнавать

---

[2] Здесь и далее я цитирую высказывания Дугласа Хофштадтера из интервью, которое я взяла у него после встречи в *Google*, причем цитаты точно отражают содержание и тон его ремарок, сделанных в присутствии инженеров *Google*.

[3] Слова Джека Шварца цит. по: G.-C. Rota, *Indiscrete Thoughts* (Boston: Berkhäuser, 1997), 22.

[4] D. R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid* (New York: Basic Books, 1979), 678. (Русское издание: Хофштадтер Д. *Гёдель. Эшер. Бах: эта бесконечная гирлянда* / Пер. с англ. М. Эскиной. – Самара: Бахрах-М, 2001.)

комбинации и распознавать абстрактные концепции шахматные программы никогда не смогут достигнуть уровня лучших шахматистов. Аргументы Хёрста казались Хофштадтеру убедительными.

Тем не менее в 1980-х и 1990-х годах компьютерные шахматные программы совершили резкий скачок в развитии, в основном за счет стремительного повышения быстродействия компьютеров. Лучшие программы по-прежнему играли не по-человечески: чтобы выбрать ход, они заглядывали далеко вперед. К середине 1990-х годов компьютер *Deep Blue*, разработанный компанией IBM специально для игры в шахматы, достиг уровня гроссмейстера. В 1997 году программа нанесла поражение действующему на тот момент чемпиону мира Гарри Каспарову в матче из шести партий. Искусство шахмат, которое еще недавно казалось вершиной человеческого разума, пало под натиском “грубой силы”.

Музыка – бастион человечности

Хотя победа *Deep Blue* подтолкнула прессу к спекуляциям о начале эпохи разумных машин, “настоящий” ИИ все еще казался делом далекого будущего. *Deep Blue* умел играть в шахматы, но не умел ничего другого. Хофштадтер ошибся насчет шахмат, но не отказывался от других возможных ответов на вопросы об искусственном интеллекте, заданные в “ГЭБ”, и особенно на первый в списке:

> **Вопрос:** Будет ли компьютер когда-нибудь сочинять прекрасную музыку?
>
> **Возможный ответ:** Да, но не скоро.

Хофштадтер продолжил:

> Музыка – это язык эмоций, и до тех пор, пока компьютеры не испытают сложных эмоций, подобных человеческим, они не смогут создать ничего прекрасного. Они смогут создавать “подделки” – поверхностные формальные имитации чужой музыки. Однако несмотря на то, что можно подумать априори, музыка – это нечто большее, чем набор синтаксических правил… Я слышал мнение, что вскоре мы сможем управлять перепрограммированной дешевой машинкой массового производства, которая, стоя у нас на столе, будет выдавать из своих стерильных внутренностей произведения, которые могли бы быть написаны Шопеном или Бахом, живи они подольше. Я считаю, что это гротескная и бессовестная недооценка глубины человеческого разума[5].

Хофштадтер называл этот ответ “одним из самых важных элементов «ГЭБ»” и готов был “поставить на него собственную жизнь”.

В середине 1990-х уверенность Хофштадтера в оценках ИИ снова пошатнулась – и довольно сильно, – когда он познакомился с программой, написанной музыкантом Дэвидом Коупом. Программа называлась *Experiments in Musical Intelligence* (“Эксперименты с музыкальным интеллектом”), или ЭМИ. Композитор и профессор музыки Коуп создал ЭМИ, чтобы она облегчила ему процесс сочинения музыки, автоматически генерируя пьесы в его характерном стиле. Но прославилась ЭМИ созданием пьес в стиле композиторов-классиков, например Баха и Шопена. ЭМИ сочиняет музыку, следуя большому набору правил, разработанных Коупом и определяющих общий синтаксис композиции. При создании новой пьесы “в стиле” конкретного композитора эти правила применяются к многочисленным примерам из его творчества.

На встрече в *Google* Хофштадтер восхищенно рассказывал о своем знакомстве с ЭМИ:

> Я сел за фортепиано и сыграл одну из мазурок ЭМИ “в стиле Шопена”. Она звучала не совсем так, как музыка Шопена, но в достаточной степени напоминала его творчество и была достаточно складной, чтобы я ощутил *глубокое* беспокойство.
>
> Музыка с детства завораживала меня и трогала до глубины души. Все мои любимые композиции кажутся мне посланием из сердца человека, который их создал. Такое впечатление, что они позволяют мне заглянуть композитору в душу. И кажется, что в мире нет *ничего* человечнее, чем выразительность музыки. Ничего. Мысль о том, что примитивная манипуляция с алгоритмами может выдавать вещи, звучащие так, словно они идут из человеческого сердца, очень и очень тревожна. Я был сражен.

[5] Ibid., 676.

Затем Хофштадтер вспомнил лекцию, которую читал в престижной Истменской школе музыки в Рочестере, в штате Нью-Йорк. Описав ЭМИ, он предложил слушателям, среди которых было несколько преподавателей композиции и теории музыки, угадать, какое из произведений, исполняемых пианистом, было (малоизвестной) мазуркой Шопена, а какое – сочинением ЭМИ. Впоследствии один из слушателей вспоминал: “Первая мазурка была изящна и очаровательна, но ей не хватало изобретательности и плавности «настоящего Шопена»… Вторая явно принадлежала Шопену – ее отличали лиричная мелодия, широкие и изящные хроматические модуляции, а также естественная и гармоничная форма”[6]. Многие преподаватели согласились с такой оценкой и шокировали Хофштадтера, проголосовав за то, что ЭМИ написала первую пьесу, а Шопен – вторую. На самом деле все было наоборот.

В переговорной *Google* Хофштадтер сделал паузу и вгляделся в наши лица. Все молчали. Наконец он продолжил: “Я испугался ЭМИ. Действительно испугался. Я возненавидел ее и узрел в ней серьезную угрозу. Она грозила уничтожить все то, что я особенно ценил в человечестве. Думаю, ЭМИ стала ярчайшим воплощением моих опасений, связанных с искусственным интеллектом”.

[6] Цит. по: D. R. Hofstadter, “Staring Emmy Straight in the Eye – and Doing My Best Not to Flinch”, in *Creativity, Cognition, and Knowledge*, ed. T. Dartnell (Westport, Conn.: Praeger, 2002), 67–100.

Google и сингулярность

Затем Хофштадтер заговорил о своем неоднозначном отношении к исследованиям ИИ, проводимым в *Google*, где занимались беспилотными автомобилями, распознаванием речи, пониманием естественного языка, машинным переводом, компьютерно-генерируемым искусством, сочинением музыки и многим другим. Тревоги Хофштадтера подкреплялись верой *Google* в Рэя Курцвейла и его представление о сингулярности, в которой ИИ, имея способность самостоятельно учиться и совершенствоваться, быстро достигнет человеческого уровня разумности, а затем и превзойдет его. Казалось, *Google* делает все возможное, чтобы ускорить этот процесс. Хотя Хофштадтер сильно сомневался в идее сингулярности, он признавал, что предсказания Курцвейла его тревожат. “Меня пугали эти сценарии. Я был настроен весьма скептически, но все равно мне казалось, что они могут оказаться правдой, даже если называемые сроки неверны. Нас застанут врасплох. Нам будет казаться, что ничего не происходит, и мы не заметим, как компьютеры вдруг станут умнее нас”.

Если это действительно произойдет, “нас отодвинут в сторону. Мы превратимся в пережитки прошлого. Мы останемся ни с чем. Может, это и правда случится, но я не хочу, чтобы это случилось *в ближайшее время*. Я не хочу, чтобы мои дети остались ни с чем”.

Хофштадтер завершил свое выступление прозрачным намеком на тех самых инженеров *Google*, которые собрались в переговорной и внимательно слушали: “Меня очень пугает, очень тревожит, очень печалит, а также ужасает, шокирует, удивляет, озадачивает и обескураживает, что люди очертя голову слепо бегут вперед, создавая все эти вещи”.

Почему Хофштадтер испуган?

Я огляделась. Казалось, слушатели озадачены и даже растеряны. Исследователей ИИ из *Google* ничто из перечисленного ничуть не пугало. Более того, это были дела давно минувших дней. Когда *Deep Blue* обыграл Каспарова, когда ЭМИ начала сочинять мазурки в стиле Шопена и когда Курцвейл написал первую книгу о сингулярности, многие из этих инженеров учились в школе и, возможно, читали "ГЭБ" и восхищались ею, несмотря на то что многие из приводимых в ней прогнозов развития ИИ к тому времени несколько устарели. Они работали в *Google* как раз для того, чтобы сделать ИИ реальностью, причем не через сотню лет, а сейчас – и как можно скорее. Они не понимали, что именно так сильно беспокоит Хофштадтера.

Люди, работающие в сфере ИИ, привыкли сталкиваться с опасениями людей со стороны, которые, вероятно, сформировали свои представления под влиянием научно-фантастических фильмов, где сверхразумные машины переходят на сторону зла. Исследователям ИИ также знакомы тревоги, что всё более сложные системы ИИ будут заменять людей в некоторых сферах, что при использовании в анализе больших данных ИИ будет нарушать конфиденциальность и потворствовать скрытой дискриминации, а также что неудачные системы ИИ, самостоятельно принимающие решения, будут сеять хаос.

Но Хофштадтера ужасало иное. Он боялся не того, что ИИ станет слишком умным, слишком вездесущим, слишком вредоносным или слишком полезным. Он боялся, что разум, творчество, эмоции и, возможно, даже сознание станут слишком *просто* воспроизводимыми, а следовательно, все те вещи, которые он особенно ценит в человечестве, окажутся не более чем "набором хитростей" и примитивный набор алгоритмов "грубой силы" сможет объяснить человеческую душу.

Как видно из "ГЭБ", Хофштадтер твердо верит, что разум и все его характеристики в полной мере проистекают из физической среды мозга и остального тела, а также из взаимодействия тела с физическим миром. Здесь нет ничего нематериального или бесплотного. Хофштадтер беспокоится о сложности. Он боится, что ИИ может показать нам, что человеческие качества, которые мы ценим больше всего, очень просто механизировать. После встречи Хофштадтер пояснил мне, рассуждая о Шопене, Бахе и других выдающихся представителях рода человеческого: "Если маленький чип сможет обесценить эти умы бесконечной остроты, сложности и эмоциональной глубины, это разрушит мое представление о сущности человечества".

Я в недоумении

За выступлением Хофштадтера последовала короткая дискуссия, в ходе которой недоумевающие слушатели побуждали Хофштадтера подробнее разъяснять опасения, связанные с ИИ и, в частности, с работой *Google*. Но коммуникационный барьер не рухнул. Встреча продолжилась презентациями и обсуждениями проектов с перерывами на кофе – все как обычно, – но о словах Хофштадтера больше никто не вспоминал. В завершение встречи Хофштадтер попросил участников поделиться своими соображениями о ближайшем будущем ИИ. Несколько исследователей из *Google* предположили, что общий ИИ человеческого уровня, вероятно, появится в течение ближайших тридцати лет, во многом благодаря успехам *Google* в совершенствовании метода "глубокого обучения", навеянного представлениями о человеческом мозге.

Я ушла со встречи в растерянности. Я знала, что Хофштадтера тревожат некоторые работы Курцвейла о сингулярности, но прежде не понимала, насколько глубоки его опасения. Я также знала, что *Google* активно занимается исследованиями ИИ, но поразилась оптимизму некоторых прогнозов о времени выхода общего ИИ на "человеческий" уровень. Лично я полагала, что ИИ достиг больших успехов в ряде узких сфер, но не приблизился к широкой, общей разумности человека – и не мог приблизиться к ней даже за столетие, не говоря уже о трех десятках лет. При этом я была уверена, что люди, имеющие другую точку зрения, серьезно недооценивают сложность человеческого разума. Я читала книги Курцвейла и по большей части находила их смехотворными. Тем не менее, прислушавшись к мнению людей, которых я уважала и которыми восхищалась, я решила критически оценить свои собственные взгляды. Возможно, полагая, что эти исследователи ИИ недооценивают людей, я сама недооценивала силу и потенциал сегодняшнего ИИ?

В последующие месяцы я начала внимательнее следить за обсуждением этих вопросов. Я вдруг стала замечать множество статей, постов и целых книг, в которых выдающиеся люди говорили нам, что прямо сейчас настает пора опасаться угроз "сверхчеловеческого" ИИ. В 2014 году физик Стивен Хокинг заявил: "Развитие полноценного искусственного интеллекта может означать конец человеческой расы"[7]. В тот же год предприниматель Илон Маск, основавший компании *Tesla* и *SpaceX*, назвал искусственный интеллект, "вероятно, величайшей угрозой нашему существованию" и сказал, что "искусственным интеллектом мы призываем демона"[8]. Сооснователь *Microsoft* Билл Гейтс согласился с ним: "В этом вопросе я согласен с Илоном Маском и другими и не понимаю, почему некоторые люди не проявляют должной озабоченности"[9]. Книга философа Ника Бострома "Искусственный интеллект" (*Superintelligence*)[10], где он рассказывает о потенциальных угрозах, которые возникнут, когда машины станут умнее людей, неожиданно стала бестселлером, несмотря на сухость и тяжеловесность повествования.

Другие ведущие мыслители выражали несогласие. Да, говорили они, нам нужно удостовериться, что программы ИИ безопасны и не могут причинить вред людям, но все сообщения о

---

[7] Цит. по: R. Cellan-Jones, "Stephen Hawking Warns Artificial Intelligence Could End Mankind", *BBC News*, Dec. 2, 2014, www.bbc.com/news/technology-30290540.

[8] M. McFarland, "Elon Musk: «With Artificial Intelligence, We Are Summoning the Demon»", *Washington Post*, Oct. 24, 2014. *https://www.washingtonpost.com/news/innovations/wp/*2014/10/24/*elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/*

[9] Bill Gates, on Reddit, Jan. 28, 2015, *www.reddit.com/r/IAmA/comments/*2*tzjp*7/*hi_reddit_im_bill_gates_and_im_back_for_my_third/*?.

[10] Бостром Н. *Искусственный интеллект: Этапы. Угрозы. Стратегии* / Пер. с англ. С. Филина. – М.: Манн, Иванов и Фербер, 2016. *(Здесь и далее в сносках, если не указано иное, – прим. перев.)*

скором появлении сверхчеловеческого ИИ серьезно преувеличены. Предприниматель и активист Митчелл Капор сказал: “Человеческий разум – удивительный, изощренный и не до конца изученный феномен. Угрозы воспроизвести его в ближайшее время нет”[11]. Робототехник (и бывший директор лаборатории ИИ в MIT) Родни Брукс согласился с ним и отметил, что мы “сильно переоцениваем способности машин – и тех, что работают сегодня, и тех, что появятся в грядущие несколько десятилетий”[12]. Психолог и исследователь ИИ Гэри Маркус даже высказал мнение, что в сфере создания “сильного ИИ” – то есть общего ИИ человеческого уровня – “прогресса почти не наблюдается”[13].

Я могла бы и дальше приводить противоречащие друг другу цитаты. В общем и целом, я пришла к выводу, что в сфере ИИ царит неразбериха. Наблюдается то ли огромный прогресс, то ли почти никакого. Появление “настоящего” ИИ ожидается то ли со дня на день, то ли через несколько веков. ИИ или решит все наши проблемы, или всех нас лишит работы, или уничтожит человеческую расу, или обесценит человечность. Стремление к нему – то ли благородная цель, то ли “призыв демона”.

[11] Цит. по: K. Anderson, “Enthusiasts and Skeptics Debate Artificial Intelligence”, *Vanity Fair*, Nov. 26, 2014.

[12] R. A. Brooks, “Mistaking Performance for Competence”, in *What to Think About Machines That Think*, ed. J. Brockman (New York: Harper Perennial, 2015), 108–111.

[13] Цит. по: G. Press, “12 Observations About Artificial Intelligence from the O’Reilly AI Conference”, *Forbes*, Oct. 31, 2016, *www.forbes.comobservations-about-artificial-intelligence-from-the-oreilly-ai-conference/sites/gilpress/*2016/10/31/12-/#886*a*6012*ea*2*e*.

О чем эта книга

Эта книга появилась, когда я предприняла попытку разобраться в истинном положении вещей в области искусственного интеллекта и понять, что компьютеры умеют сегодня и чему научатся за несколько десятков лет. Провокационные высказывания Хофштадтера на встрече в *Google* стали для меня призывом к действию, который поддержали и уверенные ответы исследователей *Google* о ближайшем будущем ИИ. В последующих главах я стараюсь прояснить, как далеко зашел искусственный интеллект, а также пролить свет на его разнообразные – и порой противоречащие друг другу – цели. Для этого я анализирую работу ряда важнейших систем ИИ, оцениваю их успехи и описываю ограничения. Я рассматриваю, насколько хорошо сегодняшние компьютеры справляются с задачами, которые, по нашему мнению, требуют интеллекта высокого уровня: как они обыгрывают людей в интеллектуальных играх, переводят тексты, отвечают на непростые вопросы и управляют транспортными средствами в сложных условиях. Я также анализирую, как они выполняют элементарные в нашем понимании задачи, которые не требуют больших интеллектуальных усилий: распознают лица и объекты на изображениях, понимают речь и текст, а также пользуются обыденным здравым смыслом.

Я пытаюсь осмыслить и более широкие вопросы, которые с самого начала подпитывали дебаты об ИИ. Что мы называем “общим человеческим” или даже “сверхчеловеческим” интеллектом? Близок ли современный ИИ к этому уровню? Можно ли сказать, что он движется в этом направлении? Какие опасности таит ИИ? Какие аспекты своего разума мы особенно ценим? В какой степени ИИ человеческого уровня заставит нас усомниться в своих представлениях о собственной человечности? Говоря словами Хофштадтера, насколько мы должны быть испуганы?

В этой книге вы не найдете общего обзора истории искусственного интеллекта, но найдете подробный анализ ряда методов ИИ, которые, возможно, уже влияют на вашу жизнь или начнут влиять на нее в скором времени, а также анализ проектов ИИ, бросающих самый смелый вызов нашему восприятию человеческой уникальности. Я хочу поделиться с вами своими соображениями и надеюсь, что вы – как и я – получите более ясное представление о том, чего уже добился ИИ и какие задачи ему предстоит решить, прежде чем машины смогут заявить о собственной человечности.

Часть I
Предыстория

Глава 1
Истоки искусственного интеллекта

Два месяца и десять мужчин в Дартмуте

Мечта о создании разумной машины – машины, которая разумна в той же степени, что и человек, или даже превосходит его, – появилась много столетий назад, но стала частью современной науки с началом эры цифровых компьютеров. Идеи, приведшие к созданию первых программируемых компьютеров, родились из стремления математиков понять человеческое мышление – в частности, логику – как механический процесс "манипуляции символами". Цифровые компьютеры, по сути, представляют собой символьные манипуляторы, которые оперируют комбинациями символов 0 и 1. Пионеры вычислительной техники, включая Алана Тьюринга и Джона фон Неймана, проводили аналогии между работой компьютеров и мозга человека, и им казалось очевидным, что человеческий разум можно воссоздать в компьютерной программе.

Большинство людей из сферы искусственного интеллекта считают, что официально она выделилась в отдельную дисциплину на маленьком семинаре, который был организован молодым математиком Джоном Маккарти и состоялся в 1956 году в Дартмуте.

В 1955 году Маккарти, которому тогда было двадцать восемь лет, поступил преподавателем на математический факультет Дартмутского колледжа. В студенческие годы он изучал психологию и новую "теорию автоматов" (которая впоследствии стала информатикой) и заинтересовался идеей о создании думающей машины. В аспирантуре на кафедре математики Принстонского университета Маккарти познакомился с Марвином Минским, который учился вместе с ним и тоже интересовался потенциалом разумных компьютеров. Окончив аспирантуру, Маккарти некоторое время работал в Лабораториях Белла и IBM, где сотрудничал с основателем теории информации Клодом Шенноном и пионером электротехники Натаниэлем Рочестером. Оказавшись в Дартмуте, Маккарти убедил Минского, Шеннона и Рочестера помочь ему с организацией "двухмесячного семинара по изучению искусственного интеллекта, который планировалось провести летом 1956 года с участием десяти человек"[14]. Термин "искусственный интеллект" был предложен Маккарти, который хотел отделить эту сферу от связанной с ней кибернетики[15]. Позже Маккарти признал, что название никому не нравилось, ведь целью был *настоящий*, а не "искусственный" интеллект, но "без названия было не обойтись", поэтому он стал использовать понятие "искусственный интеллект"[16].

Четыре организатора летнего семинара подали заявку на получение финансирования в Фонд Рокфеллера. Они написали, что в основе планируемого исследования лежит "предположение, что каждый аспект обучения и любую другую характеристику интеллекта теоретиче-

[14] J. McCarthy et al., "A Proposal for the Dartmouth Summer Research Project in Artificial Intelligence", submitted to the Rockefeller Foundation, 1955, reprinted in *AI Magazine* 27, no. 4 (2006): 12–14.

[15] Кибернетика – это междисциплинарная наука, которая изучает закономерности "управления и коммуникации в живых организмах и машинах". См. N. Wiener, *Cybernetics* (Cambridge, Mass.: MIT Press, 1961).

[16] Цит. по: N. J. Nilsson, John McCarthy: *A Biographical Memoir* (Washington, D. C.: National Academy of Sciences, 2012).

ски можно описать так точно, что можно создать машину для его моделирования”[17]. В заявке перечислялись основные темы для обсуждения – обработка естественного языка, нейронные сети, машинное обучение, абстрактные концепции и рассуждения, творческие способности, – и они до сих пор определяют сферу искусственного интеллекта.

Хотя самые мощные компьютеры в 1956 году были примерно в миллион раз медленнее современных смартфонов, Маккарти с коллегами полагали, что создание ИИ не за горами: “Мы считаем, что значительного прогресса по одной или нескольким из этих задач можно добиться, если группа правильно подобранных ученых получит возможность поработать над ними вместе в течение лета”[18].

Вскоре возникли трудности, знакомые любому, кто хоть раз пытался организовать научный семинар. Фонд Рокфеллера согласился предоставить лишь половину запрашиваемой суммы, а убедить участников приехать и остаться оказалось гораздо сложнее, чем Маккарти мог предположить. О согласии между ними не шло и речи. Было много интересных дискуссий и много противоречий. Как обычно бывает на подобных встречах, “у каждого была своя идея, большое самомнение и горячий интерес к собственному плану”[19]. Тем не менее дартмутское лето ИИ помогло добиться очень важных результатов. У области исследований появилось название. Были очерчены ее общие цели. Будущая “большая четверка” пионеров сферы – Маккарти, Минский, Аллен Ньюэлл и Герберт Саймон – встретились и построили ряд планов на будущее. По какой-то причине эти четверо после семинара смотрели в будущее искусственного интеллекта с оптимизмом. В начале 1960-х годов Маккарти основал Стэнфордский проект в области искусственного интеллекта с “целью за десять лет сконструировать полностью разумную машину”[20]. Примерно в то же время будущий нобелевский лауреат Герберт Саймон предсказал: “В ближайшие двадцать лет машины смогут выполнять любую работу, которая под силу человеку”[21]. Вскоре после этого основатель Лаборатории ИИ в MIT Марвин Минский дал прогноз, что “до смены поколений… задачи создания «искусственного интеллекта» будут в основном решены”[22].

Основные понятия и работа с ними

Пока не произошло ни одно из предсказанных событий. Насколько далеки мы от цели сконструировать “полностью разумную машину”? Потребуется ли нам для этого осуществить обратное проектирование сложнейшего человеческого мозга или же мы найдем короткий путь, где хитрого набора пока неизвестных алгоритмов будет достаточно, чтобы создать полностью разумную в нашем представлении машину? Что вообще имеется в виду под “полностью разумной машиной”?

“Определяйте понятия… или мы никогда не поймем друг друга”[23]. Этот призыв философа XVIII века Вольтера остается слабым местом дискуссий об искусственном интеллекте, поскольку главное понятие – интеллект – все еще не имеет четкого определения. Марвин Минский называл такие понятия, как “интеллект”, “мышление”, “познание”, “сознание” и “эмо-

[17] McCarthy et al., “Proposal for the Dartmouth Summer Research Project in Artificial Intelligence”.

[18] Ibid.

[19] G. Solomonoff, *Ray Solomonoff and the Dartmouth Summer Research Project in Artificial Intelligence,* 1956, accessed Dec. 4, 2018, www.raysolomonoff.com/dartmouth/dartray.pdf.

[20] H. Moravic, *Mind Children: The Future of Robot and Human Intelligence* (Cambridge, Mass.: Harvard University Press, 1988), 20.

[21] H. A. Simon, *The Shape of Automation for Men and Management* (New York: Harper & Row, 1965), 96.

[22] M. L. Minsky, *Computation: Finite and Infinite Machines* (Upper Saddle River, N. J.: Prentice-Hall, 1967), 2.

[23] B. R. Redman, *The Portable Voltaire* (New York: Penguin Books, 1977), 225.

ция", "словами-чемоданами"[24]. Каждое из них, подобно чемодану, набито кучей разных смыслов. Понятие "искусственный интеллект" наследует эту проблему, в разных контекстах означая разное.

Большинство людей согласится, что человек наделен интеллектом, а пылинка – нет. Кроме того, мы в массе своей считаем, что человек разумнее червя. Коэффициент человеческого интеллекта измеряется по единой шкале, но мы также выделяем разные аспекты интеллекта: эмоциональный, вербальный, пространственный, логический, художественный, социальный и так далее. Таким образом, интеллект может быть бинарным (человек либо умен, либо нет), может иметь диапазон (один объект может быть разумнее другого), а может быть многомерным (один человек может обладать высоким вербальным, но низким эмоциональным интеллектом). Понятие "интеллект" действительно напоминает набитый до отказа чемодан, который грозит лопнуть.

Как бы то ни было, сфера ИИ практически не принимает в расчет эти многочисленные различия. Усилия ученых направлены на решение двух задач: научной и практической. В научном направлении исследователи ИИ изучают механизмы "естественного" (то есть биологического) интеллекта, пытаясь внедрить его в компьютеры. В практическом направлении поборники ИИ просто хотят создать компьютерные программы, которые справляются с задачами не хуже или лучше людей, и не заботятся о том, *думают* ли эти программы таким же образом, как люди. Если спросить исследователей ИИ, какие цели – научные или практические – они преследуют, многие в шутку ответят, что это зависит от того, кто в данный момент финансирует их работу.

В недавнем отчете о текущем состоянии ИИ комитет видных исследователей определил эту область как "раздел информатики, изучающий свойства интеллекта посредством синтеза интеллекта"[25]. Да, определение закольцовано. Впрочем, тот же комитет признал, что определить область сложно, но это и к лучшему: "Отсутствие точного, универсального определения ИИ, вероятно, помогает области все быстрее расти, развиваться и совершенствоваться"[26]. Более того, комитет отмечает: "Практики, исследователи и разработчики ИИ ориентируются по наитию и руководствуются принципом «бери и делай»".

Анархия методов

Участники Дартмутского семинара 1956 года озвучивали разные мнения о правильном подходе к разработке ИИ. Одни – в основном математики – считали, что языком рационального мышления следует считать математическую логику и дедуктивный метод. Другие выступали за использование индуктивного метода, в рамках которого программы извлекают статистические сведения из данных и используют вероятности при работе с неопределенностью. Третьи твердо верили, что нужно черпать вдохновение из биологии и психологии и создавать программы по модели мозга. Как ни странно, споры между сторонниками разных подходов не утихают по сей день. Для каждого подхода было разработано собственное множество принципов и техник, поддерживаемых отраслевыми конференциями и журналами, но узкие специальности почти не взаимодействуют между собой. В недавнем исследовании ИИ отмечается: "Поскольку мы не имеем глубокого понимания интеллекта и не знаем, как создать общий ИИ, чтобы идти по пути настоящего прогресса, нам нужно не закрывать некоторые направления исследований, а принимать «анархию методов», царящую в сфере ИИ"[27].

---

[24] M. L. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind* (New York: Simon & Schuster, 2006), 95.

[25] *One Hundred Year Study on Artificial Intelligence* (AI100), 2016 Report, 13, ai100.stanford.edu/2016-report.

[26] Ibid., 12.

[27] J. Lehman, J. Clune and S. Risi, "An Anarchy of Methods: Current Trends in How Intelligence Is Abstracted in AI", *IEEE*

Однако с 2010-х годов одно семейство ИИ-методов, в совокупности именуемых глубоким обучением (или глубокими нейронными сетями), выделилось из анархии и стало господствующей парадигмой ИИ. Многие популярные медиа сегодня ставят знак равенства между понятиями "искусственный интеллект" и "глубокое обучение". При этом они совершают досадную ошибку, и мне стоит прояснить различия между терминами. ИИ – это область, включающая широкий спектр подходов, цель которых заключается в создании наделенных интеллектом машин. Глубокое обучение – лишь один из этих подходов. Глубокое обучение – лишь один из множества методов в области "машинного обучения", подобласти ИИ, где машины "учатся" на основе данных или собственного "опыта". Чтобы лучше понять эти различия, нужно разобраться в философском расколе, который произошел на заре исследования ИИ, когда произошло разделение так называемых символического и субсимволического ИИ.

Символический ИИ

Давайте сначала рассмотрим *символический* ИИ. Программа символического ИИ знает слова или фразы ("символы"), как правило понятные человеку, а также правила комбинирования и обработки этих символов для выполнения поставленной перед ней задачи.

Приведу пример. Одной ранней программе ИИ присвоили громкое имя "Универсальный решатель задач" (*General Problem Solver*, или GPS)[28]. (Прошу прощения за сбивающую с толку аббревиатуру: Универсальный решатель задач появился раньше системы глобального позиционирования, ныне известной как GPS.) УРЗ мог решать такие задачи, как задача о миссионерах и людоедах, над которой вы, возможно, ломали голову в детстве. В этой известной задаче три миссионера и три людоеда должны переправиться через реку на лодке, способной выдержать не более двух человек. Если на одном берегу окажется больше (голодных) людоедов, чем (аппетитных) миссионеров, то… думаю, вы поняли, что произойдет. Как всем шестерым переправиться на другой берег без потерь?

Создатели УРЗ, когнитивисты Герберт Саймон и Аллен Ньюэлл, записали, как несколько студентов "размышляют вслух", решая эту и другие логические задачи. Затем Саймон и Ньюэлл сконструировали программу таким образом, чтобы она копировала ход рассуждений студентов, который ученые признали их мыслительным процессом.

Я не буду подробно описывать механизм работы УРЗ, но его символическую природу можно разглядеть в формулировке программных инструкций. Чтобы поставить задачу, человек писал для УРЗ подобный код:

```
ТЕКУЩЕЕ СОСТОЯНИЕ:
ЛЕВЫЙ-БЕРЕГ = [3 МИССИОНЕРА, 3 ЛЮДОЕДА, 1 ЛОДКА]
ПРАВЫЙ-БЕРЕГ = [ПУСТО]

ЖЕЛАЕМОЕ СОСТОЯНИЕ:
ЛЕВЫЙ-БЕРЕГ = [ПУСТО]
ПРАВЫЙ-БЕРЕГ = [3 МИССИОНЕРА, 3 ЛЮДОЕДА, 1 ЛОДКА]
```

Если говорить обычным языком, эта инструкция показывает, что изначально левый берег реки "содержит" трех миссионеров, трех людоедов и одну лодку, в то время как правый не содержит ничего. Желаемое состояние определяет цель программы – переправить всех на правый берег реки.

На каждом шаге программы УРЗ пытается изменить текущее состояние, чтобы сделать его более похожим на желаемое состояние. В этом коде у программы есть "операторы" (в

Intelligent Systems 29, no. 6 (2014): 56–62.

[28] A. Newell and H. A. Simon, "GPS: A Program That Simulates Human Thought", P-2257, Rand Corporation, Santa Monica, Calif. (1961).

форме подпрограмм), которые могут преобразовывать текущее состояние в новое состояние, и "правила", кодирующие ограничения задачи. Например, один оператор перемещает некоторое количество миссионеров и людоедов с одного берега реки на другой:

```
ПЕРЕМЕСТИТЬ (#МИССИОНЕРОВ, #ЛЮДОЕДОВ, С-БЕРЕГА, НА-БЕРЕГ)
```

Слова в скобках называются параметрами, и после запуска программа заменяет эти слова на числа или другие слова. Параметр **#миссионеров** заменяется на количество перемещаемых миссионеров, параметр **#людоедов** заменяется на количество перемещаемых людоедов, а параметры **с-берега** и **на-берег** заменяются на **"левый-берег"** и "**правый-берег**" в зависимости от того, с какого берега нужно переместить миссионеров и людоедов. В программе закодировано знание, что лодка перемещается вместе с миссионерами и людоедами.

Прежде чем программа сумеет применить этот оператор с конкретными значениями параметров, она должна свериться с закодированными правилами: так, за один раз можно переместить не более двух человек, а если в результате применения оператора на одном берегу окажется больше людоедов, чем миссионеров, применять его нельзя.

Хотя эти символы обозначают знакомые людям понятия – "миссионеры", "людоеды", "лодка", "левый берег", – запускающий программу компьютер, конечно, не понимает значения символов. Можно заменить параметр "**миссионеры**" на "**z372b**" или любой другой бессмысленный набор знаков, и программа будет работать точно так же. Отчасти поэтому она называется *Универсальным* решателем задач. Компьютер определяет "значение" символов в зависимости от того, как их можно комбинировать и соотносить друг с другом и как ими можно оперировать.

Адвокаты символического подхода к ИИ утверждали, что невозможно наделить компьютер разумом, не написав программы, копирующие человеческий мозг. Они полагали, что для создания общего интеллекта необходима лишь верная программа обработки символов. Да, эта программа работала бы гораздо сложнее, чем в примере с миссионерами и людоедами, но все равно состояла бы из символов, комбинаций символов, правил и операций с символами. В итоге символический ИИ, примером которого стал Универсальный решатель задач, доминировал в сфере ИИ в первые три десятилетия ее существования. Самым заметным его воплощением стали *экспертные системы*, которые использовали созданные людьми правила для компьютерных программ, чтобы выполнять такие задачи, как постановка медицинских диагнозов и принятие юридических решений. Символический ИИ по-прежнему применяется в нескольких сферах – я приведу примеры таких систем позже, в частности при обсуждении подходов ИИ к построению логических выводов и здравому смыслу.

Субсимволический ИИ: перцептроны

Вдохновением для символического ИИ послужила математическая логика и описание людьми своих сознательных мыслительных процессов. *Субсимволический* ИИ, напротив, вдохновлялся нейробиологией и стремился ухватить порой бессознательные мыслительные процессы, лежащие в основе так называемых процессов быстрого восприятия, например распознавания лиц или произносимых слов. Программы субсимволического ИИ не содержат понятного людям языка, который мы наблюдали в примере с миссионерами и людоедами. По сути, они представляют собой набор уравнений – настоящие дебри непонятных операций с числами. Как я поясню чуть дальше, такие системы на основе данных учатся выполнять поставленную перед ними задачу.

Одной из первых субсимволических ИИ-программ, созданных по модели мозга, стал перцептрон, изобретенный в конце 1950-х годов психологом Фрэнком Розенблаттом[29]. Сегодня

[29] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological*

термин “перцептрон” кажется заимствованным из научной фантастики пятидесятых годов (как мы увидим, вскоре за ним последовали “когнитрон” и “неокогнитрон”), но перцептрон стал важной вехой развития ИИ и может считаться авторитетным прадедом самого успешного инструмента современного ИИ, глубоких нейронных сетей.

Розенблатт изобрел перцептрон, обратив внимание на то, как нейроны обрабатывают информацию. Нейрон – это клетка мозга, которая получает электрический или химический импульс от связанных с нею нейронов. Грубо говоря, нейрон суммирует все импульсы, которые получает от других нейронов, и сам посылает импульс, если итоговая сумма превышает определенный порог. Важно, что разные связи (*синапсы*) конкретного нейрона с другими нейронами имеют разную силу, а потому, суммируя импульсы, нейрон придает больше веса импульсам от сильных связей, чем импульсам от слабых связей. Нейробиологи полагают, что поправки на силу связей между нейронами – важнейший элемент процесса обучения, происходящего в мозге.

С точки зрения специалиста по информатике (или, как в случае с Розенблаттом, психолога), обработку информации нейронами можно смоделировать в компьютерной программе – перцептроне, – которая преобразует много численных входных сигналов в один выходной сигнал. Аналогия между нейроном и перцептроном показана на рис. 1. На рис. 1A мы видим нейрон с ветвистыми дендритами (волокнами, которые проводят входящие импульсы в клетку), телом клетки и аксоном (или выводным каналом). На рис. 1B изображен простой перцептрон. Как и нейрон, перцептрон суммирует все входящие сигналы. Если итоговая сумма равняется *порогу* перцептрона или превышает его, перцептрон выдает значение 1 (“передает сигнал”); в противном случае он выдает значение 0 (“не передает сигнал”). Чтобы смоделировать различную силу связей нейрона, Розенблатт предложил присваивать каждому входному сигналу перцептрона численный *вес* и умножать входной сигнал на его вес, прежде чем прибавлять к сумме. *Порог* перцептрона – это число, определяемое программистом (или, как мы увидим, узнаваемое самим перцептроном).

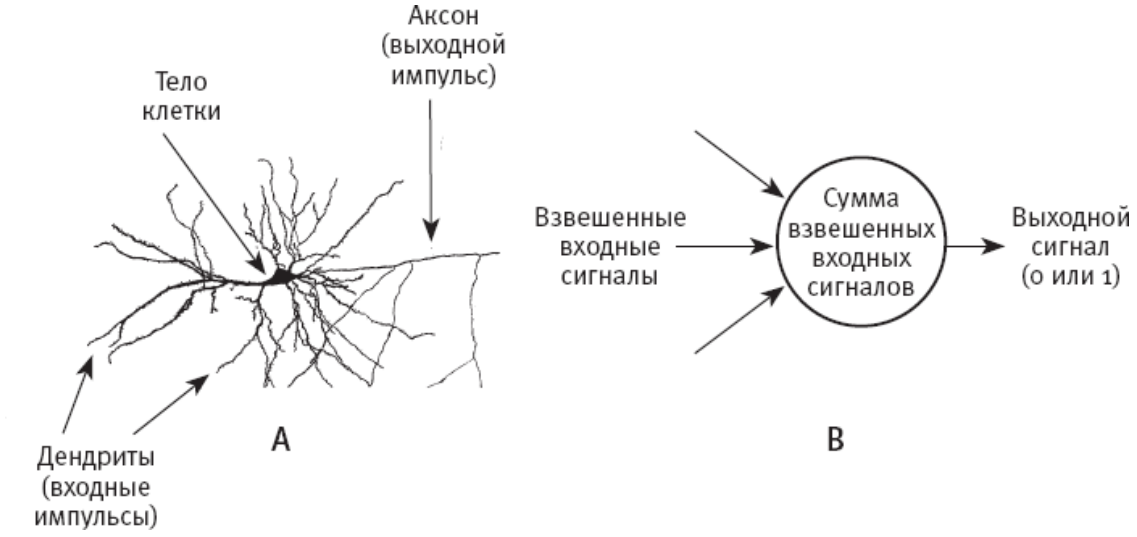

Рис. 1. A: нейрон в мозге; B: простой перцептон

Иными словами, перцептрон – это простая программа, которая принимает решение “да или нет” (1 или 0) в зависимости от того, достигает ли сумма взвешенных входных сигналов порогового значения. Вероятно, вы тоже время от времени принимаете такие решения в жизни. Например, вы узнаете мнение нескольких друзей о конкретном фильме, но вкусам одних дру-

Review 65, no. 6 (1958): 386–408.

зей доверяете больше, чем вкусам других. Если сумма “дружеских восторгов” – при большем весе мнений тех друзей, которым вы доверяете больше, – достаточно высока (то есть превышает некоторый неосознанный порог), вы решаете посмотреть фильм. Именно так перцептрон выбирал бы фильмы к просмотру, если бы у него были друзья.

Вдохновленный сетями нейронов в мозге, Розенблатт предположил, что сети перцептронов могут выполнять визуальные задачи, например справляться с распознаванием объектов и лиц. Чтобы понять, как это может работать, давайте изучим, как с помощью перцептрона решить конкретную визуальную задачу: распознать рукописные цифры вроде тех, что показаны на рис. 2.

Давайте сделаем перцептрон детектором восьмерок – в таком случае он будет выдавать единицу, если входным сигналом служит изображение цифры 8, и ноль, если на входном изображении будет любая другая цифра. Чтобы создать такой детектор, нам нужно (1) понять, как превратить изображение в набор численных входных сигналов, и (2) определить численные значения весов и порог перцептрона для формирования верного выходного сигнала (1 для восьмерок и 0 для других цифр). Я рассмотрю эту задачу более подробно, поскольку многие из этих принципов понадобятся нам при обсуждении нейронных сетей и их применения в компьютерном зрении.

Рис. 2. Примеры рукописных цифр

Входные сигналы нашего перцептрона

На рис. 3А показана увеличенная рукописная восьмерка. Каждый квадрат координатной сетки – это пиксель с численным значением “насыщенности”: насыщенность белых квадратов равняется 0, насыщенность черных – 1, а насыщенность серых имеет промежуточное значение. Допустим, все изображения, которые мы даем перцептрону, подогнаны к единому размеру – 18 × 18 пикселей. На рис. 3В показан перцептрон для распознавания восьмерок. У этого перцептрона 324 (то есть 18 × 18) входных сигнала, каждый из которых соответствует одному пикселю из сетки 18 ×18. При получении изображения, подобного показанному на рисунке 3А,

каждый входной сигнал настраивается на насыщенность соответствующего пикселя. Каждому входному сигналу также присваивается свой вес (на рисунке не показан).

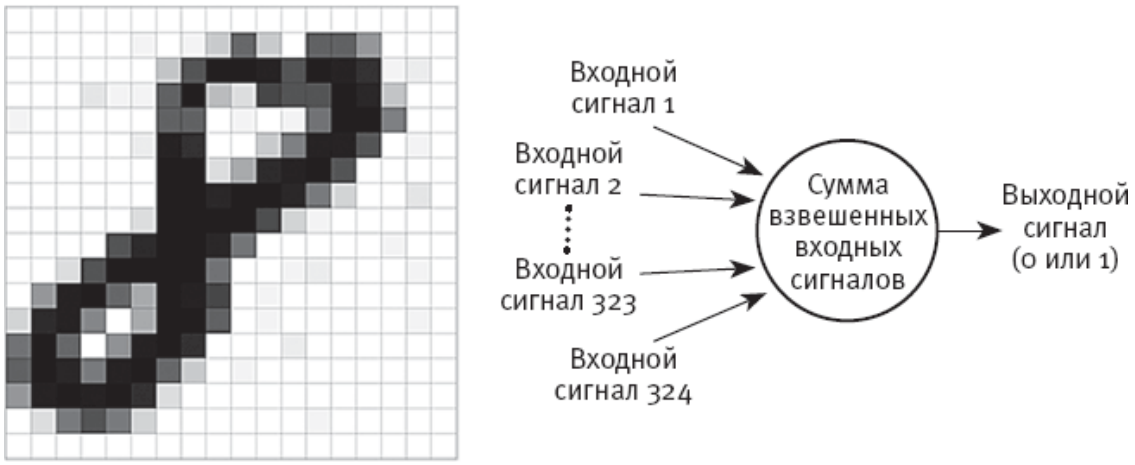

Рис. 3. Иллюстрация перцептрона, который распознает рукописные восьмерки. Каждый пиксель на изображении 18 × 18 пикселей соответствует одному входному сигналу перцептрона, что дает 324 (= 18 × 18) входных сигнала.

Как узнать веса и порог перцептрона

В отличие от символической системы Универсального решателя задач, которую я описала ранее, перцептрон не имеет очевидных правил для выполнения задачи, а все его "знания" закодированы в числах, определяющих веса входных сигналов и пороговое значение. В ряде статей Розенблатт показал, что при корректных весах и пороговом значении такой перцептрон, как на рисунке 3В, вполне неплохо справляется с такими задачами на восприятие, как распознавание простых рукописных цифр. Но как именно определить корректные веса и пороговое значение для конкретной задачи? И снова Розенблатт предложил ответ, навеянный работой мозга: перцептрон должен сам *узнавать* эти значения. Но каким образом? Вторя популярным в то время теориям бихевиоральной психологии, Розенблатт считал, что перцептроны должны обучаться, накапливая *условный рефлекс*. Отчасти вдохновленный работой бихевиориста Б. Ф. Скиннера, который обучал крыс и голубей выполнять задачи с помощью положительного и отрицательного подкрепления, Розенблатт полагал, что перцептрон следует *обучать* на примерах: его нужно вознаграждать, когда он выдает верный результат, и наказывать, когда он ошибается. Теперь такая форма обучения в ИИ называется обучением с учителем. В ходе обучения система получает пример и генерирует выходной сигнал, а затем получает "сигнал от учителя", который показывает, насколько выходной сигнал системы отличается от верного. Затем система использует этот сигнал, чтобы скорректировать веса и пороговое значение.

Концепция обучения с учителем – ключевой элемент современного ИИ, поэтому ее стоит разобрать подробнее. Как правило, обучение с учителем требует большого набора положительных (скажем, коллекции восьмерок, написанных разными людьми) и отрицательных (скажем, коллекции других рукописных цифр, среди которых нет восьмерок) примеров. Каждый пример *размечается* человеком, который присваивает ему определенную категорию (метку) – здесь это "восьмерка" и "не восьмерка". Метка применяется в качестве контрольного сигнала. Некоторые положительные и отрицательные примеры используются для *тренировки* системы и формируют *тренировочное множество*. Оставшиеся примеры – *тестовое множество* – используются для оценки работы системы после обучения, чтобы понять, насколько хорошо она научилась правильно отвечать на запросы в целом, а не только на обучающие примеры.

Вероятно, самым важным в информатике стоит признать понятие “алгоритм”. Оно обозначает “рецепт” со списком шагов, которые компьютер может предпринять для решения конкретной задачи. Главным вкладом Фрэнка Розенблатта в ИИ стало создание особого алгоритма, названного алгоритмом обучения перцептрона. С помощью этого алгоритма перцептрон можно научить на примерах определять веса и пороговое значение для получения верных ответов. Вот как он работает: сначала весам и порогу присваиваются случайные значения в диапазоне от –1 до 1. В нашем примере первому входному сигналу может быть присвоен вес 0,2, второму – вес –0,6 и так далее. Пороговым значением может стать 0,7. С генерацией начальных значений без труда справится компьютерная программа, называемая генератором случайных чисел.

Теперь мы можем приступать к процессу обучения. Перцептрон получает первый обучающий пример, не видя метку с верной категорией. Перцептрон умножает каждый входной сигнал на его вес, суммирует результаты, сравнивает сумму с пороговым значением и выдает либо 1, либо 0. Здесь выходной сигнал 1 означает, что перцептрон распознал восьмерку, а выходной сигнал 0 – что он распознал “не восьмерку”. Далее в процессе обучения выходной сигнал перцептрона сравнивается с верным ответом, который дает присвоенная человеком метка (“восьмерка” или “не восьмерка”). Если перцептрон прав, веса и пороговое значение не меняются. Если же перцептрон ошибся, веса и пороговое значение слегка корректируются так, чтобы сумма входных сигналов в этом тренировочном примере оказалась ближе к нужной для верного ответа. Более того, степень изменения каждого веса зависит от соответствующего значения входного сигнала, то есть вина за ошибку в основном возлагается на входные сигналы, которые сильнее других повлияли на результат. Например, в восьмерке на рис. 3А главным образом на результат повлияли бы более насыщенные (здесь – черные) пиксели, в то время как пиксели с нулевой насыщенностью (здесь – белые) не оказали бы на него никакого влияния. (Для любопытных читателей я описала некоторые математические подробности в примечании[30].)

Все шаги повторяются на каждом из обучающих примеров. Процесс обучения много раз проходится по всем обучающим примерам, слегка корректируя веса и пороговое значение при каждой ошибке перцептрона. Обучая голубей, психолог Б. Ф. Скиннер обнаружил, что учиться лучше постепенно, совершая множество попыток, и здесь дело обстоит точно так же: если слишком сильно изменить веса и пороговое значение после одной попытки, система может научиться неправильному правилу (например, чрезмерному обобщению, что “нижняя и верхняя половины восьмерки всегда равны по размеру”). После множества повторов каждого обучающего примера система (как мы надеемся) окончательно определяет набор весов и пороговое значение, при которых перцептрон дает верные ответы для всех обучающих примеров. На этом этапе мы можем проверить перцептрон на примерах из тестового множества и увидеть, как он справляется с распознаванием изображений, не входивших в обучающий набор.

Детектор восьмерок полезен, когда вас интересуют только восьмерки. Но что насчет распознавания других цифр? Не составляет труда расширить перцептрон таким образом, чтобы он выдавал десять выходных сигналов, по одному на каждую цифру. Получая пример рукописной цифры, перцептрон будет выдавать единицу в качестве выходного сигнала, соответствую-

[30] Математически алгоритм обучения перцептрона описывается следующим образом. Для каждого веса w_j: $w_j \leftarrow w_j + \eta\,(t - y)\,x_j$, где t – верный выходной сигнал (1 или 0) для заданного входного сигнала, y – фактический выходной сигнал перцептрона, x_j – входной сигнал, связанный с весом w_j, а η – *скорость обучения*, задаваемая программистом. Стрелка обозначает обновление. Порог учитывается путем создания дополнительного “входного сигнала” x_0 с постоянным значением 1, которому присваивается вес $w_0 = -$*порог*. При наличии этого дополнительного входного сигнала и веса (называемого смещением) перцептрон дает сигнал на выходе, только если сумма входных сигналов, помноженных на веса (то есть скалярное произведение входного вектора и вектора веса) больше или равняется 0. Часто входные значения масштабируются и подвергаются другим преобразованиям, чтобы веса не становились слишком велики.

щего этой цифре. При наличии достаточного количества примеров расширенный перцептрон сможет узнать все необходимые веса и пороговые значения, используя алгоритм обучения.

Розенблатт и другие исследователи показали, что сети перцептронов можно научить выполнять относительно простые задачи на восприятие, а еще Розенблатт математически доказал, что теоретически достаточно обученные перцептроны могут безошибочно выполнять задачи определенного, хотя и строго ограниченного класса. При этом было непонятно, насколько хорошо перцептроны справляются с более общими задачами ИИ. Казалось, эта неопределенность не мешала Розенблатту и его спонсорам из Научно-исследовательского управления ВМС США делать до смешного оптимистичные прогнозы о будущем алгоритма. Освещая пресс-конференцию Розенблатта, состоявшуюся в июле 1958 года, газета *The New York Times* написала:

> Сегодня ВМС продемонстрировали зародыш электронного компьютера, который, как ожидается, сможет ходить, говорить, видеть, писать, воспроизводить себя и сознавать свое существование. Было сказано, что в будущем перцептроны смогут узнавать людей, называть их по именам и мгновенно переводить устную речь и тексты с одного языка на другой[31].

Да, даже в самом начале ИИ страдал от шумихи. Вскоре я расскажу о печальных последствиях такого ажиотажа. Но пока позвольте мне на примере перцептронов объяснить основные различия между символическим и субсимволическим подходом к ИИ.

Поскольку “знания” перцептрона состоят из набора чисел, а именно – определенных в ходе обучения весов и порогового значения, – сложно выявить правила, которые перцептрон использует при выполнении задачи распознавания. Правила перцептрона не символические: в отличие от символов Универсального решателя задач, таких как **ЛЕВЫЙ-БЕРЕГ, #МИССИОНЕРОВ** и **ПЕРЕМЕСТИТЬ**, веса и порог перцептрона не соответствуют конкретным понятиям. Довольно сложно преобразовать эти числа в понятные людям правила. Ситуация существенно усложняется в современных нейронных сетях с миллионами весов.

Можно провести грубую аналогию между перцептронами и человеческим мозгом. Если бы я могла заглянуть к вам в голову и понаблюдать за тем, как некоторое подмножество ста миллиардов ваших нейронов испускает импульсы, скорее всего, я бы не поняла, ни о чем вы думаете, ни какие “правила” применяете при принятии конкретного решения. Тем не менее человеческий мозг породил язык, который позволяет вам использовать символы (слова и фразы), чтобы сообщать мне – часто недостаточно четко, – о чем вы думаете и почему приходите к определенным выводам. В этом смысле наши нервные импульсы можно считать *субсимволическими*, поскольку они лежат в основе символов, которые каким-то образом создает наш мозг. Перцептроны, а также более сложные сети искусственных нейронов, называются “субсимволическими” по аналогии с мозгом. Их поборники считают, что для создания искусственного интеллекта языкоподобные символы и правила их обработки должны не программироваться непосредственно, как для Универсального решателя задач, а рождаться в нейроноподобных архитектурах точно так же, как интеллектуальная обработка символов рождается в мозге.

Ограниченность перцептронов

После Дартмутского семинара 1956 года доминирующее положение в сфере ИИ занял символический лагерь. В начале 1960-х годов, пока Розенблатт увлеченно работал над перцептроном, большая четверка “основателей” ИИ, преданных символическому лагерю, создала

[31] Цит. по: M. Olazaran, “A Sociological Study of the Official History of the Perceptrons Controversy”, *Social Studies of Science* 26, no. 3 (1996): 611–659.

авторитетные – и прекрасно финансируемые – лаборатории ИИ: Марвин Минский открыл свою в MIT, Джон Маккарти – в Стэнфорде, а Герберт Саймон и Аллен Ньюэлл – в Университете Карнеги – Меллона. (Примечательно, что эти университеты по сей день входят в число самых престижных мест для изучения ИИ.) Минский, в частности, полагал, что моделирование мозга, которым занимался Розенблатт, ведет в тупик и ворует деньги у более перспективных проектов символического ИИ[32]. В 1969 году Минский и его коллега по MIT Сеймур Пейперт опубликовали книгу "Перцептроны"[33], в которой математически доказали, что существует крайне ограниченное количество типов задач, поддающихся *безошибочному* решению перцептроном, а алгоритм обучения перцептрона не сможет показывать хорошие результаты, когда задачи будут требовать большого числа весов и порогов.

Минский и Пейперт отметили, что если перцептрон усовершенствовать, добавив дополнительный "слой" искусственных нейронов, то количество типов задач, которые сможет решать устройство, значительно возрастет[34]. Перцептрон с таким дополнительным слоем называется многослойной нейронной сетью. Такие сети составляют основу значительной части современного ИИ, и я подробно опишу их в следующей главе. Пока же я отмечу, что в то время, когда Минский и Пейперт писали свою книгу, многослойные нейронные сети еще не были широко изучены, в основном потому что не существовало общего алгоритма, аналогичного алгоритму обучения перцептрона, для определения весов и пороговых значений.

Ограниченность простых перцептронов, установленная Минским и Пейпертом, была уже известна людям, работавшим в этой сфере[35]. Сам Фрэнк Розенблатт много работал с многослойными перцептронами и признавал, что их сложно обучать[36]. Но последний гвоздь в крышку гроба перцептронов вогнала не математика Минского и Пейперта, а их рассуждения о многослойных нейронных сетях:

> [Перцептрон] обладает многими свойствами, привлекающими внимание: линейность, интригующая способность к обучению, очевидная простота перцептрона как разновидности устройства для параллельных вычислений. Нет никаких оснований предполагать, что любое из этих достоинств распространяется на многослойный вариант. Тем не менее мы считаем важной исследовательской задачей разъяснить (или отвергнуть) наше интуитивное заключение о том, что обсуждаемое расширение бесплодно[37].

Ой-ой! Сегодня последнее предложение этого отрывка, возможно, сочли бы "пассивно-агрессивным". Такие негативные спекуляции отчасти объясняют, почему в конце 1960-х финансирование исследований нейронных сетей прекратилось, хотя государство продолжало вливать немалые деньги в символический ИИ. В 1971 году Фрэнк Розенблатт утонул в возрасте сорока трех лет. Лишившись главного идеолога и большей части государственного финансирования, исследования перцептронов и других систем субсимволического ИИ практически остановились. Ими продолжали заниматься лишь несколько отдельных академических групп.

---

[32] M. A. Boden, *Mind as Machine: A History of Cognitive Science* (Oxford: Oxford University Press, 2006), 2:913.

[33] M. L. Minsky and S. L. Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, Mass.: MIT Press, 1969). (Минский М., Пейперт С. *Персептроны* / Пер. с англ. Г. Гимельфарба и В. Шарыпанова – М.: Издательство "Мир", 1971.)

[34] Выражаясь техническим языком, любую булеву функцию можно вычислить с помощью полностью подключенной многослойной сети с линейными пороговыми значениями и одним внутренним ("скрытым") слоем.

[35] Olazaran, "Sociological Study of the Official History of the Perceptrons Controversy".

[36] G. Nagy, "Neural Networks – Then and Now", *IEEE Transactions on Neural Networks* 2, no. 2 (1991): 316–318.

[37] Minsky and Papert, "Perceptrons", 231–232. (Пер. с англ. Г. Гимельфарба и В. Шарыпанова.)

Зима ИИ

Тем временем поборники символического ИИ писали заявки на гранты, обещая скорые прорывы в таких областях, как понимание речи и языка, построение логических выводов на основе здравого смысла, навигация роботов и беспилотные автомобили. К середине 1970-х годов были успешно развернуты некоторые узкие экспертные системы, но обещанных прорывов общего характера так и не произошло.

Это не укрылось от внимания финансирующих организаций. Британский Совет по научным исследованиям и Министерство обороны США подготовили отчеты, в которых дали крайне отрицательную оценку прогрессу и перспективам исследований ИИ. В частности, в британском отчете отмечалось, что некоторые надежды вселяет продвижение в области специализированных экспертных систем – "программ, написанных для работы в узких сферах, где программирование полностью принимает во внимание человеческий опыт и человеческие знания в соответствующей области", – но подчеркивалось, что текущие результаты работы "над программами общего назначения, ориентированными на копирование механизма решения широкого спектра задач с человеческого [мозга], удручают. Вожделенная долгосрочная цель исследований в сфере ИИ кажется все такой же далекой"[38]

[38] J. Lighthill, "Artificial Intelligence: A General Survey", in *Artificial Intelligence: A Paper Symposium* (London: Science Research Council, 1973).

Конец ознакомительного фрагмента.