СОДЕРЖАНИЕ

Предисловие · 11

Введение · 15

1. Начала (1956–1970) · 21
   1.1. Зарождение ИИ · 21
   1.2. Программа Артура Сэмюэла для игры в шашки · 25
   1.3. ИИ около 1968 года · 29
   1.4. Символы и Logic Theorist · 30
   1.5. Распознавание сцен · 33
   1.6. Распространение ограничений и поиск с возвратом · 36
   1.7. Обработка естественного языка и SHRDLU · 38
   1.8. Обучение в раннем ИИ · 46
   1.9. Перцептроны · 48

2. Рассуждение и представление знаний (1970–1985) · 56
   2.1. Рассуждение, решение задач и планирование · 56
   2.2. STRIPS · 58
   2.3. Упорядочивание действий · 61
   2.4. Эвристики планирования · 63
   2.5. Экспертные системы · 65
   2.6. Представление знаний · 70
   2.7. Рассуждения и гипотеза о физической символьной системе · 76

3. Рассуждения в условиях неопределенности (1980–1990) · 79
   3.1. Экспертные системы и неопределенность · 79
   3.2. Частотная и байесовская статистика · 82
   3.3. Теория вероятностей · 85
   3.4. Неудовлетворенность стандартной теорией вероятностей · 93

3.5. Альтернативы стандартной теории вероятностей · 95
3.6. Байесовские сети · 98

4. Шахматы (1965–1997) · 107
4.1. Организация шахматной программы · 109
4.2. Шахматная программа Бернштейна · 112
4.3. Mac Hack и Chess 4.0 · 113
4.4. Deep Blue · 115

5. Компьютерное зрение (1970–2000) · 117
5.1. Хьюбел и Визель · 117
5.2. Свертка · 119
5.3. Стереоскопическое зрение · 122
5.4. Визуальные признаки · 127
5.5. Обучение зрению · 132
5.6. Модели деформируемых частей · 135

6. Распознавание речи (1971–1985) · 138
6.1. От звука к речи · 138
6.2. Распознавание речи и модель зашумленного канала · 144
6.3. Языковые модели и статистика · 147
6.4. Скрытые марковские модели · 154

7. Обучение языку (1985–2010) · 158
7.1. Ранний машинный перевод · 159
7.2. Параллельные корпуса · 160
7.3. Извлечение словарей для перевода · 162
7.4. Снова о языковых моделях · 166
7.5. Графические модели и грамматика · 169

8. Глубокое обучение (1989–2016) · 179
8.1. Глубокое обучение и компьютерное зрение · 183
8.2. Состязательные примеры · 191
8.3. Графические процессоры и аппаратное обеспечение для ИИ · 194
8.4. Распределенное представление слов · 199
8.5. Рекуррентные нейронные сети · 205
8.6. Автокодировщики и генеративные модели · 208

9. Обучение с подкреплением и игра го (1990–2017) · 211
9.1. Го · 211
9.2. Нейросети для го · 213
9.3. Обучение с подкреплением · 214
9.4. AlphaGo, продолжение · 222
9.5. AlphaGo Zero · 223
9.6. Игры со случайным элементом · 225
10. Масштабное обучение (2017–2023) · 228
10.1. Все, что вам нужно, — это внимание · 228
10.2. Большие языковые модели · 235
10.3. Тест Тьюринга · 242
10.4. DALL-E: от слов к изображениям · 245
10.5. AlphaFold · 251
10.6. ChatGPT и GPT-4 · 259
11. ИИ: настоящее и будущее · 263
Хронология ИИ · 276

MIT Press выражает искреннюю благодарность
Майклу Литтману за помощь
в завершении рукописи профессора Черняка
после его безвременной кончины в 2023 году.

Моим коллегам.
ИИ — это *сложно*.

Предисловие

В 1992 ГОДУ я пришел в кабинет профессора Юджина Черняка как будущий аспирант. Последние четыре года я работал младшим сотрудником в научной группе, которая принципиально избегала термина «искусственный интеллект» — какникак, еще не закончилась вторая «зима ИИ». Вместо этого моя группа сосредоточилась на *статистической обработке естественного языка*, где значения слов извлекались не из вручную проработанных структур данных, а из анализа их совместной встречаемости (коллокаций) в тысячах и десятках тысяч примеров из необработанных, естественно-языковых текстов. Я встретился с Юджином Черняком, признанным авторитетом в области обработки естественного языка, чтобы спросить, интересно ли ему будет научное сотрудничество со мной в этой сфере, если я выберу для учебы Брауновский университет.

По большому счету, он ответил мне отказом. Он предпочитал заниматься *вероятностной* обработкой естественного языка. В рамках этого подхода представления смысла, разработанные вручную и преобладавшие на заре этой области, не отбрасывались, а, наоборот, дополнялись числовой информацией. Это помогало системам концентрироваться на наиболее *вероятных* интерпретациях, а не просто на правдоподобных.

Я прежде не слышал о таком различии и, честно говоря, был настроен немного скептически. Я не ви-

дел, каким образом вероятностная или статистическая обработка естественного языка могли бы привести к появлению систем общего искусственного интеллекта. Однако статистический подход, по крайней мере, обходил проблему «бутылочного горлышка» ручного труда и обещал возможность масштабирования для работы с текстами из реального мира.

Я ушел со встречи воодушевленным, ведь у меня состоялась стимулирующая беседа с таким выдающимся исследователем в области ИИ. Я не чувствовал, что на меня смотрят свысока или пренебрегают моими идеями, а, наоборот, ощутил себя участником широкой дискуссии, где рассматривались действительно глубокие и захватывающие проблемы. Уверен, читая эту книгу, вы почувствуете, как это было — обсуждать с ним эти вопросы. Он подходил к задаче превращения компьютеров в компетентных пользователей языка практически со всех мыслимых сторон. У него сформировались чрезвычайно обоснованные мнения, но он всегда был открыт к рассмотрению нового подхода, если тот был ему незнаком. В нем чувствовалось это отношение: «Интересно! Давайте продумаем это вместе». Это сделало его феноменальным специалистом в своей области, вдохновляющим профессором и интеллектуальным примером для подражания для многих из нас. Может, и правда, в ИИ нет ничего такого уж зазорного.

В итоге год спустя я стал аспирантом в Брауновском университете. Мои исследования были сосредоточены в области обучения с подкреплением, но между языком и принятием решений существовала поразительная общая структура: обе сферы по своей сути имеют дело с *последовательностями* — слов в одном случае и действий в другом — и способами их генерации. Я решил заглянуть к Юджину

и был ошеломлен, обнаружив, что в своих исследованиях он теперь целиком и полностью погрузился в статистическую обработку естественного языка. Более того, он вел курс по этой теме и писал учебник, чтобы помочь студентам разобраться в фундаментальных концепциях этой области. Он практически в одночасье превратился из скептика в эксперта мирового уровня, потому что осознал, что статистические методы — это более многообещающий путь к созданию программ, способных эффективно использовать язык. Со временем его работы по статистическому синтаксическому анализу, или парсингу, — аннотированию последовательности слов с указанием их взаимосвязей для распознавания глубинного смысла — стали лучшими в мире. Вместе со студентами он применял статистическое обучение для создания программного обеспечения, которое могло «читать» предложения из газетных статей и парсить их на уровне, почти не уступающем людям-аннотаторам.

В итоге я окончил аспирантуру в Брауне, защитив докторскую, и последним курсом, который я брал, был курс по учебнику Юджина! Я вернулся туда лет двадцать спустя уже как преподаватель. Таким образом, я был коллегой Юджина, когда область обработки естественного языка пережила свой самый недавний переход — от статистической обработки к нейронным языковым моделям. Как и до этого со статистическими методами, Юджин начал преподавать курс по глубокому обучению и написал по нему учебник. Он полностью принял эту новую перспективу и ее обещание решить проблемы, с которыми он боролся десятилетиями, если не всю свою карьеру. Его последняя опубликованная статья «Парсинг как языковое моделирование», написанная с его последним аспирантом До Кук Чхве (ДК),

показала, как его любимая задача парсинга может быть решена в связке с новейшими достижениями в обучении нейронных сетей. В конечном счете, диссертация ДК продемонстрировала, как достичь производительности, сопоставимой с человеческой, при парсинге коллекции газетных статей, позволив Юджину и ДК заявить, что их работа раз и навсегда решила эту проблему.

Юджин никогда не видел в парсинге конечную цель ИИ, но он был твердо убежден, что это важный промежуточный этап. Он исследовал эту проблему, последовательно применяя логические, вероятностные, статистические и, наконец, нейросетевые подходы, пока она в конечном счете не была решена. Его главный тезис в этой книге состоит в том, что искусственный интеллект по-настоящему может начаться только сейчас. Думаю, вы получите удовольствие, читая, как он выстраивает свою аргументацию, делясь остроумием и проницательностью, которые он приобрел, будучи непосредственным наблюдателем и участником более чем семидесятилетней истории этой области.

Теперь, когда ИИ совершает взлет, я скажу от имени своих коллег по области: спасибо, Юджин, за то, что помогли нам отрываться от земли. Теперь для нас нет преград!

Майкл Л. Литтман, январь 2024 года

Введение

Области искусственного интеллекта (ИИ) около шестидесяти пяти лет, но она уже играет исключительно важную роль в интеллектуальной жизни нашего времени. ИИ вносит весомый (возможно, ключевой) вклад в современную теорию сознания, и на то есть веские основания. Его также называют критически важной технологией для нашего дальнейшего процветания, хотя я думаю, что это утверждение несколько преувеличено. При этом ИИ часто ставят в один ряд с квантовыми компьютерами, значение которых *очень сильно* раздуто. (Но это тема для другой книги.)

Эта книга — моя попытка написать интеллектуальную историю ИИ как единого целого. Я говорю «интеллектуальная история», потому что это рассказ о ключевых идеях, которые легли в основу этой дисциплины. Обычная «история ИИ» содержала бы больше дат и имен, была бы энциклопедичной, освещала бы такие темы, как основные организации (крупнейшая — Ассоциация содействия развитию искусственного интеллекта, AAAI), финансирование или ее место в компьютерных науках (до недавнего времени и то и другое было на низком уровне), и содержала бы биографии ключевых фигур в этой области.

Интеллектуальная история ИИ интересна сама по себе. Она показывает, что у ИИ с самого начала была верная цель — написание программ, демон-

стрирующих интеллектуальное поведение, — но пути ее достижения были совершенно ошибочными. В частности, у нас было два фундаментальных заблуждения: первое — что обучение является проблемой, а не решением. Если бы на заре этой области вы спросили исследователя ИИ о ключевых проблемных направлениях, он бы перечислил: компьютерное зрение, логический вывод, представление знаний, понимание языка… и обучение. Никому бы не пришло в голову создавать программу-переводчик, которая сама бы училась переводить, — точно так же, как никто не стал бы строить ракету, которая сама бы училась лететь на Луну. В такой постановке это кажется очевидным, но это неверно.

Нашей второй большой ошибкой была наша приверженность гипотезе физических символьных систем:

> Физическая символьная система располагает необходимыми и достаточными средствами для общего интеллектуального действия.

«Символ» здесь — это знак, который представляет нечто иное, подобно тому, как музыкальная нота представляет конкретный звук, а слово — идею или объект. «Физическая символьная система» — это компьютер, который оперирует символами. В те времена, когда в это верили практически все, авторство этой идеи приписывали Аллену Ньюэллу и Герберту Саймону, возглавлявшим исследования ИИ в Университете Карнеги—Меллона. Ньюэлл и Саймон действительно первыми сформулировали это в явной форме, но, поскольку в моем повествовании это подается, скажем так, как большая ошибка, спешу добавить, что в то время в это верил практически *каждый* — и я в том числе.

Новый подход был прекрасно описан Джеффри Хинтоном, который в моем повествовании выступает в роли крестного отца нового ИИ на основе глубокого обучения:

> Теперь мы мыслим внутренние представления как огромные векторы и не рассматриваем логику как парадигму того, как заставить все работать. Мы просто считаем, что можно иметь эти огромные нейронные сети, которые обучаются, и поэтому, вместо того чтобы программировать, вы просто заставляете их всему учиться самим. (https://www.forbes.com/sites/peterhigh/2016/06/20/deep-learning-pioneer-geoff-hinton-helps-shape-googles-drive-to-put-ai-everywhere/)

Что и подводит нас ко второй причине — сейчас подходящий момент для интеллектуальной истории ИИ. Автор этих строк полностью разделяет новые взгляды. И хотя я убежден, что логика того, как мы пришли к нынешнему положению дел, является мощным аргументом в их пользу, эту точку зрения разделяют не все. Более того, издательство этой книги, MIT Press, опубликовавшее первый крупный учебник по глубокому обучению[1], выпустило затем как минимум три книги, в которых утверждается, что роль нового ИИ в лучшем случае преувеличена, а в худшем — он движется в неверном направлении[2].

---

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016; Ян Гудфеллоу, Иошуа Бенджио, Аарон Курвилль. *Глубокое обучение*. Москва: ДМК Пресс, 2018.

2. Ronald J. Brachman and Hector J. Levesque. Toward a new science of common sense. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 12245–12249, 2022; Hector J. Levesque. *Common Sense, the Turing Test, and the Quest for Real AI*. Cam-

Возвращаясь к этой книге, ее название, «ИИ и я», — это сигнал читателю, что перед ним авторский взгляд. Всю свою интеллектуальную жизнь я был практиком в области ИИ, и именно мой жизненный путь в этой сфере привел меня к тем выводам, к которым я пришел. Однако должен предупредить: я не историк науки и у меня нет того целостного видения области, которым обладал бы ученый-историк, взявшийся за написание исчерпывающей истории. Мои преимущества — это, скорее, мой возраст, который обеспечил мне долгий путь в ИИ, и мое умение собирать слова в предложения, а предложения — в абзацы.

Поскольку это авторский взгляд, мой рассказ имеет определенные недостатки, главный из которых — я включил в него слишком мало интересных идей, открытых моими коллегами. Поступить иначе — обобщить все из области ИИ — было бы работой всей жизни. Посвящение в книге — слабая компенсация за огромный массив этих упущений. Эта книга также американоцентрична, так как я смотрел на всю эту историю со своей колокольни в США. Приношу свои извинения французам, японцам, британцам и другим.

Эта книга рассчитана на широкого, но образованного читателя. Однако во многих случаях я освещаю идеи более детально, чем это принято в подобных книгах. Я делаю это отчасти потому, что могу, или, по крайней мере, надеюсь, что могу. История ИИ

---

bridge, MA: MIT Press, 2017; Gary Marcus and Ernest Davis. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon, 2019; Гэри Маркус и Эрнест Дэвис, *Искусственный интеллект: перезагрузка. Как создать машинный разум, которому действительно можно доверять*. Москва: Альпина ПРО, 2022.

до сего дня, к сожалению, — это история попыток исследователей ухватиться за решение очень сложной проблемы: как создать разумный артефакт из неразумных частей. Лишь в последние пятнадцать лет мы наблюдаем нечто похожее на то, что можно назвать кумулятивным прогрессом, и, углубляясь в детали, я хочу убедить вас, что мы, исследователи ИИ, не тратили время впустую. Идеи, которые нам пришлось отвергнуть или хотя бы отложить, были разумными, но не продуктивными. Но чтобы увидеть это и, следовательно, лучше оценить наше текущее, куда более удачное положение, необходимо понять, что именно раньше шло не так. Для этого мне и нужна дополнительная детализация. Если результат покажется вам непонятным, дайте мне знать, и в следующий раз я постараюсь лучше.

Пока вы здесь, я должен сделать некоторые пояснения относительно того, как устроена эта книга. Как и положено историческому труду, я начинаю с начала, а последние главы посвящены завершающему этапу. С серединой все менее ясно. Когда в области всего несколько практиков, она будет единой хотя бы потому, что, даже при наличии отдельных субдисциплин, у вас слишком мало коллег, на которых можно ориентироваться. Единение в конце — это уже настоящая конвергенция областей благодаря поразительной применимости нейронных сетей. Однако довольно скоро в истории ИИ субдисциплины разошлись и, например, исследователи в области компьютерного зрения имели мало общего с теми, кто интересовался играми. Поэтому начиная со второй главы главы посвящены отдельным субдисциплинам, а временной порядок в основном соблюдается в рамках самих глав. В порядке глав тоже есть своя логика. Например, хотя исследования в области компьютерного зрения начали формироваться в се-

редине 1970-х, только в 1980-х они выработали уникальный стиль (по крайней мере, по моему мнению). Поэтому в начале глав читатели часто будут обнаруживать, что переносятся в прошлое. Чтобы внести ясность, я последовал совету одного из рецензентов и добавил даты в названия глав.

Несколько моих коллег оказали мне ценную помощь и поддержку. Эрни Дэвис поделился своими примечательными и обширными знаниями во многих областях нашей науки. Питер Норвиг был первым, кто после меня прочел рукопись целиком, и, сумев разглядеть в ней ту книгу, которую я и задумал написать, очень меня воодушевил. Рао Камбхампати помог мне увидеть ИИ с точки зрения исследователей из субдисциплины ИИ, известной как «планирование», хотя он, вероятно, не согласится с моим взглядом на вещи. И наконец, мой студент Дэвид Макклоски провел на удивление внимательную вычитку книги и помог мне стереть следы моей неспособности замечать опечатки.

Я уверен, что, если бы не мое намерение обеспечить единство взгляда, не прибегая к помощи со стороны, вышеприведенный список был бы намного длиннее, а книга, возможно, значительно лучше. Но это была бы уже другая книга.

Провиденс, Род-Айленд
Май 2023 года