ПРЕДИСЛОВИЕ

Период существования биологического разума – лишь краткий промежуток между ранними формами жизни и долгой эрой машин.

Мартин Рис. Из интервью сайту The Conversation, апрель 2017.

Искусственный интеллект и вокруг него

Многие новейшие технологии ИИ находят повсеместное применение, часто даже не называясь искусственным интеллектом: как только что-то становится достаточно полезным и распространенным, его перестают называть ИИ.

Ник Бостром. ИИ готовится превзойти возможности человеческого мозга. CNN.com, 2006

На протяжении всей истории человечества тайны разума, природа мышления и возможность создания искусственных существ будоражили воображение художников, ученых, философов и даже богословов. Мифология, изобразительное искусство, музыка и литература полны образов и историй, связанных с *автоматонами* – подвижными механическими устройствами, созданными в подражание живым существам. Наше увлечение искусственным интеллектом (ИИ) – то есть разумным поведением машин – также выражается в жутких и сверхъестественных сюжетах блокбастеров и видеоигр, где появляются роботы, наделенные эмоциями, и существа с совершенным интеллектом, непостижимым для человека.

В этой книге мы отправимся в долгое путешествие во времени от древних игр к современным вычислительным методам, основанным на искусственных нейросетях, которые учатся и совершенствуются, зачастую не требуя – или почти не требуя – целевого программирования и заданных правил. На этом пути нам встретятся диковинные чудеса, например таинственные медные рыцари из цикла легенд о короле Артуре, а также творение французского изобретателя Жака де Вокансона – гиперреалистичная утка-автоматон, которая спустя более 250 лет вдохновила американского писателя Томаса Пинчона на создание одного из персонажей романа «Мейсон и Диксон». Мы познакомимся с каталонским философом XIII в. Раймундом Луллием, который одним из первых системно подошел к вопросу об искусственной генерации идей с помощью механического устройства. Мы перенесемся в 1893 г. и посмотрим на забавного черного страуса, созданного Электрическим Бобом; в этой истории, как и в романе «Паровой человек в прериях», отразился особый интерес ко всему механическому в Викторианскую эпоху – этакий викторианский стимпанк.

Ближе к нашему времени мы встретимся с Артуром Сэмюэлом из *IBM*, который в 1952 г. создал одну из первых компьютерных программ для игры в шашки, а в 1955 г. – программу, которая *научилась* играть в эту игру без посторонней помощи. Сегодня термином «искусственный интеллект» часто обозначают системы, созданные для обучения, решения проблем и взаимодействия с людьми посредством

обработки естественного языка. Умные помощники, такие как Алекса от *Amazon*, Сири от *Apple* и Кортана от *Microsoft*, обладают некоторыми возможностями ИИ.

В этой книге мы также рассмотрим непростые этические вопросы, связанные с использованием ИИ, и даже следующую проблему: как поместить сложные системы ИИ – если они станут слишком разумными и опасными – в герметичные ящики, чтобы изолировать их от внешнего мира? Конечно, границы и масштабы ИИ со временем меняются, и некоторые специалисты предлагают более общие определения, под которые подпадает целый ряд технологий, помогающих людям выполнять мыслительные операции. Поэтому, чтобы шире осветить историю ИИ, я включил в книгу несколько машин и механизмов, которые помогают решать задачи, требующие умственных усилий и подсчетов. Среди таких устройств – счеты, Антикитерский механизм (125 до н. э.), ЭНИАК (1946) и другие изобретения. В конце концов, без этих простейших технологий у нас не было бы современных шахматных программ и систем управления транспортом.

Читая эту книгу, помните: даже если какие-то гипотезы из прошлого или предсказания по поводу искусственных существ кажутся нам неправдоподобными, любая давняя идея внезапно может воплотиться в жизнь, если для этого найдется достаточно быстрый и совершенный компьютер. Наши технологические прогнозы – и даже наши легенды – по меньшей мере представляют собой увлекательную картину познания и творчества и показывают, как мы проникаем в разные культуры и эпохи, чтобы понять друг друга и выяснить, что ценно и сакрально для нашего общества. Но, отдавая дань человеческим воображению и изобретательности, крайне важно задумываться о нежелательных последствиях, в том числе о потенциальной опасности ИИ. В 2014 г. физик-теоретик Стивен Хокинг сказал в интервью Би-би-си: «Развитие полноценного искусственного разума может положить конец человеческой расе… Этот разум возьмет инициативу на себя и станет сам себя совершенствовать со все возрастающей скоростью». Иными словами, существует вероятность, что объекты с ИИ станут настолько разумными и умелыми, что, постоянно улучшая себя, в конце концов создадут некий сверхразум, потенциально чрезвычайно опасный для человечества. Этот сценарий стремительного технологического роста, иногда называемый *технологической сингулярностью*, может привести к невообразимым изменениям в цивилизации, обществе и жизни людей.

Таким образом, несмотря на многочисленные потенциальные преимущества ИИ – беспилотные автомобили, эффективные бизнес-процессы и даже помощь роботов-компаньонов в самых разных делах, – необходимо проявлять особую осторожность при разработке автономных комплексов вооружения и не слишком полагаться на технологии ИИ с порой непостижимыми механизмами. Например, исследования показывают, что некоторые системы распознавания образов на основе ИИ (нейросети) можно легко «обмануть» и заставить ошибочно идентифицировать животных как винтовки или принять самолет за собаку. Для этого достаточно немного изменить изображения таким образом, что люди даже ничего не заподозрят. Если террористу удастся сделать торговый центр или больницу похожими на военную цель для беспилотника, последствия могут

быть ужасными. С другой стороны, вполне возможно, что боевые машины с настроенными датчиками и заданными этическими правилами могли бы снизить число жертв среди мирного населения. Чтобы потенциальная опасность ИИ-устройств не перечеркивала их ценные преимущества, в этой сфере необходимо создать продуманную нормативную базу.

Мы все больше полагаемся на ИИ с его многочисленными сложными нейросетями глубокого обучения, и одновременно с этим развивается одна интересная область: разработка систем ИИ, которые смогут *объяснить* человеку, каким образом они пришли к тому или иному решению. Однако, заставив ИИ объяснять самого себя, мы тем самым ограничим его возможности – по крайней мере, в некоторых случаях. Дело в том, что многие из этих систем способны создавать гораздо более сложные модели реальности, чем люди могут себе представить. Эксперт по ИИ Дэвид Ганнинг даже предполагает, что самая высокопроизводительная система окажется и самой труднообъяснимой.

Структура и цель этой книги

Меня давно увлекает вычислительная техника и интересуют проблемы, возникающие на переднем крае науки. В этой книге, адресованной широкой аудитории, я предлагаю краткий путеводитель по любопытным и вместе с тем важным практическим идеям из истории *искусственного интеллекта* – хотя сам этот термин был предложен только в 1955 г. информатиком Джоном Маккарти. Каждая глава состоит всего из нескольких абзацев, так что книгу можно читать с любого места, не продираясь через многословные описания. Конечно, такой формат не позволил мне углубляться в подробности, однако в разделе «Примечания и список литературы» я предлагаю материалы для дальнейшего чтения и поиска источников цитат или трудов упомянутых авторов.

Главы этой книги охватывают такие области, как философия, поп-культура, информатика, социология и теология, а также темы, которые интересуют меня лично. В молодости я был очарован книгой Ясии Рейхардт «Кибернетическая проницательность: компьютер и искусство», опубликованной в 1968 г. В книге рассказывалось, как компьютеры создают стихи, картины, музыку и многое другое. Также меня поражает, каких успехов в области искусства достигли специалисты по ИИ, используя порождающие состязательные сети для создания потрясающих фотореалистичных изображений смоделированных лиц, цветов и птиц. Порождающие состязательные сети – это две нейросети, которые «натравлены» друг на друга: одна генерирует идеи и паттерны, а другая оценивает результаты.

Сегодня возможности применения ИИ кажутся безграничными, и в разработки в этой области ежегодно вкладываются миллиарды долларов. Технологии ИИ использовали, например, для расшифровки документов из Ватиканского секретного архива: ученые пытались разобрать сложные рукописные тексты из этой огромной исторической коллекции. ИИ также помогает прогнозировать землетрясения, интерпретировать медицинские снимки, распознавать речь и предсказывать время смерти пациента на основе информации из его электронной медицинской карты. С помощью ИИ придумывают шутки, игры и головоломки, формулируют

математические теоремы, создают патентуемые изобретения, разрабатывают инновационные конструкции антенн, новые оттенки красок, парфюмерные ароматы и многое другое. Уже сегодня многие из нас разговаривают со своими смартфонами и прочими устройствами, а в будущем наши отношения с машинами станут еще более близкими и похожими на человеческие.

Главы расположены в хронологическом порядке с указанием года, связанного с важным событием, книгой или открытием. Датировка часто условна, некоторые годы приводятся приблизительно; там, где это было возможно, я попытался объяснить, почему указал ту или иную дату.

Как легко заметить, больше половины глав приходится на период после 1950 г. Даниэль Кревье, автор книги «Бурная история поиска искусственного интеллекта» (1993), отмечает, что в 1960-е гг. «искусственный интеллект расцвел тысячей цветов. Специалисты по ИИ использовали новые методы программирования для решения многих проблем, которые, хоть и были реальными, оказались значительно упрощены – отчасти ради разделения задач, требующих решения, отчасти для того, чтобы втиснуть их в крошечную память компьютеров того времени».

Тайны сознания, недостатки ИИ и природа разума будут изучаться еще многие годы; но эти проблемы интересовали людей с древних времен. Памела Маккордак в своей книге «Машины, которые думают» высказывает предположение, что ИИ начался с желания древних людей «выковать богов».

Грядущие открытия, связанные с ИИ, войдут в число величайших достижений человечества. История ИИ – это история не только о том, как мы создаем свое будущее, но и о том, как люди будут жить в условиях бурного развития интеллекта и творческих возможностей. Что будет вкладываться в понятие «человек» через сто лет? Каким будет общество, в котором повсеместно станут использоваться устройства с ИИ? Что произойдет с рабочими местами? Будем ли мы влюбляться в роботов?

Если методы и модели ИИ уже помогают решать, кого нанять на работу, с кем пойти на свидание, кто получит условно-досрочное освобождение, кто более склонен к психическим расстройствам и как управлять беспилотными автомобилями и дронами, то какой уровень контроля над нашей жизнью мы доверим системам ИИ будущего? Если они все чаще принимают решения за нас, легко ли будет обмануть какой-нибудь модуль ИИ и заставить его совершить катастрофическую ошибку? Как специалистам по ИИ выяснить, почему одни алгоритмы и архитектуры машинного обучения эффективнее других, и в то же время упростить воспроизведение чужих экспериментов и их результатов?

Наконец, есть ли гарантии, что устройства на основе ИИ будут действовать этично или у машин когда-либо появятся психические состояния и чувства, свойственные людям? Несомненно, устройства с ИИ станут чем-то вроде протезов для нашего слабого мозга и помогут нам мыслить и мечтать по-новому. Для меня искусственный интеллект – источник постоянного удивления по поводу границ разума, будущего человечества и нашего места в огромном пространственно-временном ландшафте, который мы называем своим домом.

КРЕСТИКИ-НОЛИКИ

По данным археологов, нечто похожее на игру с выстраиванием трех элементов в ряд существовало еще примерно в 1300 г. до н. э. в Древнем Египте. При игре в крестики-нолики два игрока по очереди вписывают свои символы (О или Х) в клетки поля размером 3 × 3. Выигрывает тот, кто первым проставит три своих знака в ряд по горизонтали, вертикали или диагонали.

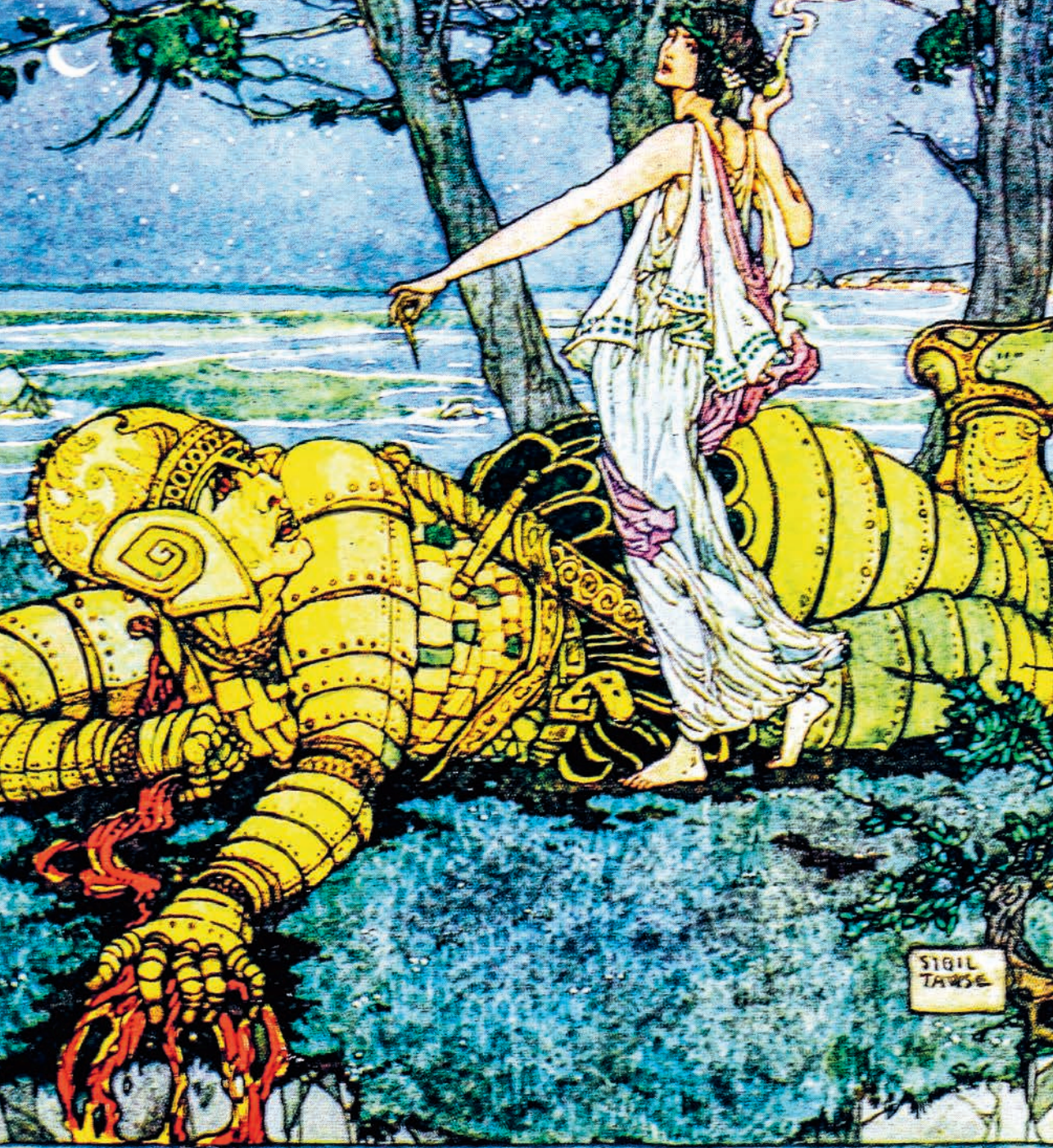
Крестики-нолики попали в эту книгу потому, что их часто упоминают при объяснении базовых принципов программирования и искусственного интеллекта из-за простоты их игровых деревьев (где узлы графа – это позиции в игре, а ребра – ходы). Крестики-нолики – это так называемая игра с полной информацией, поскольку все игроки в курсе всех сделанных ходов. Кроме того, это последовательная игра без рандомизации: игроки ходят по очереди и не используют игральные кости.

Крестики-нолики можно назвать атомом, на основе которого веками формировались молекулы более сложных позиционных игр. Даже при минимальных вариациях и расширениях эта простая игра становится труднейшей задачей, решение которой требует большого количества времени. Математики и любители головоломок усложняли крестики-нолики, добавляя дополнительные клетки и измерения, а также необычные игровые поверхности, например прямоугольные или квадратные поля, соединенные по краям в форме тора (бублика) или бутылки Клейна (поверхности, у которой только одна сторона).

Рассмотрим некоторые любопытные особенности этой игры. Всего существует 362 880 (9!, то есть 1 × 2 × 3 × 4 × ... × 9) возможных сценариев заполнения поля двумя игроками. Однако, если рассматривать все возможные партии, при которых игра заканчивается в 5, 6, 7, 8 или 9 ходов, наберется 255 168 таких партий. В 1960 г. ИИ-система *MENACE* (хитроумная конструкция из спичечных коробков и разноцветных шариков) научилась играть в крестики-нолики путем обучения с подкреплением. В начале 1980-х г. компьютерные гении Дэнни Хиллис и Брайан Сильверман с несколькими друзьями сконструировали из 10 тысяч деталей конструктора *Tinkertoy*® компьютер, который играл в крестики-нолики. В 1998 г. ученые и студенты Университета Торонто создали робота для игры в трехмерные крестики-нолики (4 × 4 × 4) с человеком.

СМ. ТАКЖЕ Мельница Лейбница (1714), Обучение с подкреплением (1951), Четыре в ряд (1988), Реверси (1997), Решение для игры вари (2002)

Крестики-нолики можно сделать более сложными для людей и машин с ИИ, расширив стандартное поле 3 × 3 до больших размеров, добавив новые измерения и эффект гравитации, при котором каждый элемент опускается в нижнюю свободную позицию, например как в этой трехмерной версии 4 × 4 × 4.



MEDEIA AND TALVS

ТАЛОС

«Многим людям образ Талоса знаком по его воплощению в виде бронзового гиганта в фильме 1963 г. „Ясон и Аргонавты“, – пишет Брайан Хотон. – Но откуда взялась идея Талоса и мог ли он быть первым роботом в истории?»

Согласно греческой мифологии, Талос был огромным бронзовым автоматоном («роботом»), созданным для защиты Европы – матери критского царя Миноса – от захватчиков, пиратов и других врагов. Он был запрограммирован патрулировать берега острова и трижды в день обходил по кругу весь Крит. Порой, чтобы остановить неприятелей, он бросал в них огромные валуны. В других случаях этот гигантский робот прыгал в огонь, раскалялся докрасна, а затем обхватывал тело врага и сжигал его заживо. Иногда Талоса изображали в виде крылатого существа – как на монетах из критского города Феста, датируемых приблизительно 300 г. до н. э. Изображения Талоса также были обнаружены на вазах, созданных около 400 г. до н. э.

Существуют разные версии сотворения и гибели Талоса. В одном мифе его по просьбе Зевса создал Гефест – греческий бог огня и обработки металлов, покровитель кузнецов и других ремесленников. Поскольку Талос был автоматоном, его внутренняя структура по сложности уступала человеческой; по сути, у Талоса имелась одна-единственная вена, которая тянулась от шеи к лодыжке. Снизу вена была запечатана и защищена от протечки бронзовым гвоздем. По одной из легенд, колдунья Медея свела Талоса с ума при помощи духов смерти (их называли «керами») и заставила выбить гвоздь. Божественная кровь (ихор) хлынула у него из лодыжки, «как расплавленный свинец», и великан умер.

Талос – лишь один из примеров того, как древние греки представляли себе роботов и самодвижущиеся автоматы. Здесь также стоит упомянуть труды математика Архита Тарентского (428–347 до н. э.), который, возможно, придумал и создал механического летающего голубя, приводимого в движение паром.

СМ. ТАКЖЕ Водяные часы Ктесибия (ок. 250 до н. э.), Медные рыцари из легенды о Ланселоте (ок. 1220), Голем (1580), «Франкенштейн» (1818)

Изображение Талоса из «Историй о богах и героях» Томаса Булфинча (1920), выполненное английской художницей Сибил Таус (1886–1971).

«ОРГАНОН» АРИСТОТЕЛЯ

Греческий философ Аристотель (384–322 до н. э.) затрагивал в своих работах несколько важных тем, которые и по сей день интересуют исследователей ИИ. В своей книге «Политика» Аристотель высказал предположение, что когда-нибудь автоматы заменят рабов: «Если бы каждое орудие могло выполнять свойственную ему работу само, по данному ему приказанию или даже его предвосхищая, и уподоблялось бы статуям Дедала или треножникам Гефеста, о которых Гомер говорит, что они “сами собой входили в собрание богов”, если бы ткацкие челноки сами ткали, а плектры сами играли на кифаре, тогда и зодчие не нуждались бы в работниках, а господам не нужны были бы рабы»[1].

Аристотель также положил начало системному изучению логики. В своих трудах под общим названием «Органон» (*др.-греч.* «инструмент», «метод») он предлагает приемы выяснения истины и осмысления мира. Основной инструмент в арсенале Аристотеля – *силлогизм*, трехступенчатый аргумент, например: «Все женщины смертны; Клеопатра – женщина; следовательно, Клеопатра смертна». Если две предпосылки истинны, то и заключение должно быть истинным. Аристотель также проводит различие между частностями и универсалиями (то есть общими категориями). Например, *Клеопатра* – это частное понятие, тогда как *женщина* и *смертны* – универсальные. Когда речь идет об универсалиях, им предшествуют слова *все*, *некоторые* или *ни один*. Аристотель проанализировал множество возможных типов силлогизмов и показал, какие из них состоятельны.

Аристотель также анализировал силлогизмы с модальной логикой – то есть утверждения, содержащие слова *возможно* или *обязательно*. Современная математическая логика далеко ушла от аристотелевской методологии, а его приемы были доработаны для применения к суждениям с другой структурой, включая те, что выражают более сложные отношения, и те, что содержат более одного квантора, как, например, фраза «Ни одному человеку не нравятся все люди, которым не нравятся некоторые люди». И все же глубокие изыскания Аристотеля в области логики считаются одним из величайших достижений человечества, давшим толчок многим разработкам в области математики и искусственного интеллекта.

СМ. ТАКЖЕ Талос (ок. 400 до н. э.), Булева алгебра (1854), Нечеткая логика (1965)

1 Пер. С. А. Жебелева.

Этот впечатляющий бюст Аристотеля – римская копия бронзового оригинала работы древнегреческого скульптора Лисиппа, жившего в IV в. до н. э.