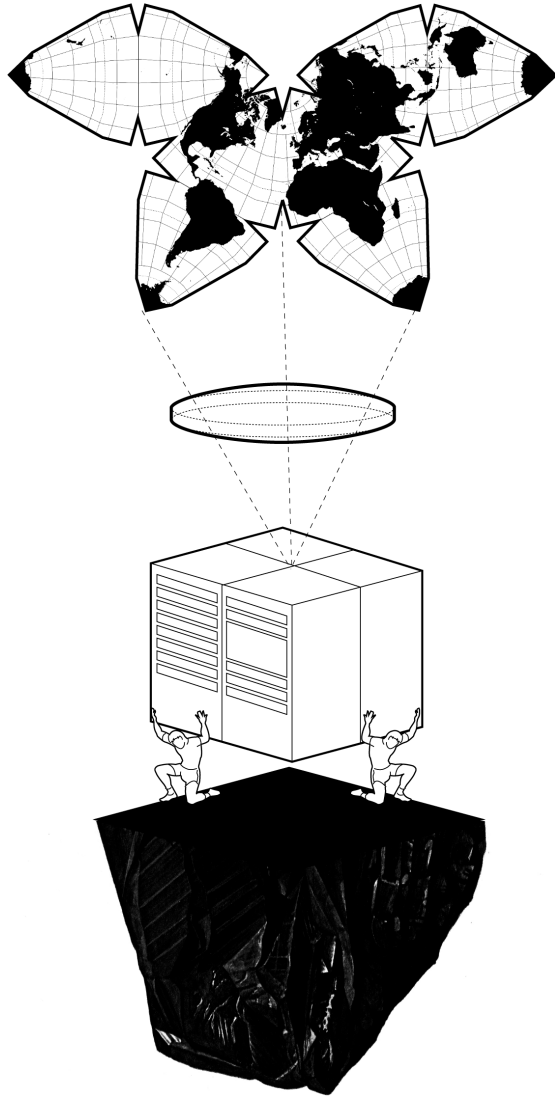


KATE CRAWFORD

ATLAS OF AI

POWER, POLITICS
AND THE PLANETARY COSTS
OF ARTIFICIAL INTELLIGENCE



КЕЙТ КРОУФОРД

**АТЛАС
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА**

РУКОВОДСТВО ДЛЯ БУДУЩЕГО

Издательство АСТ

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

5

ГЛАВА 1. ЗЕМЛЯ

25

ГЛАВА 2. ТРУД

53

ГЛАВА 3. ДАННЫЕ

85

ГЛАВА 4. КЛАССИФИКАЦИЯ

117

ГЛАВА 5. ЭМОЦИИ

143

ГЛАВА 6. ГОСУДАРСТВО

171

ЗАКЛЮЧЕНИЕ

199

ЭПИЛОГ

217

БЛАГОДАРНОСТИ

227

ПРИМЕЧАНИЯ

232

БИБЛИОГРАФИЯ

262

ВВЕДЕНИЕ

Самая умная лошадь в мире

В конце девятнадцатого века Европу покорила лошадь по кличке Ганс. Умный Ганс поражал воображение: он решал математические задачи, определял время, дни в календаре, различал музыкальные тона, составлял слова и предложения. Люди стекались посмотреть, как немецкий жеребец отстукивает копытом ответы на сложные задачи и неизменно приходит к правильному ответу. «Сколько будет два плюс три?» Ганс старательно отстукивал копытом по земле пять раз. «Какой сегодня день недели?» Лошадь стучала копытом, верно указывая на каждую букву на специально изготовленной доске. Ганс освоил даже более сложные вопросы, например, такие: «Я загадал число. Я вычитаю девять и получаю три. Какое число я загадал?» К 1904 году Умный Ганс стал международной знаменитостью, а газета *New York Times* назвала его «Чудесной берлинской лошадью, способной на все, кроме разговоров». [1]

Дрессировщик Ганса, отставной учитель математики по имени Вильгельм фон Остен, давно увлекался интеллектом животных. Фон Остен пытался обучать котят и медвежат порядковым номерам, однако успеха он добился лишь после того, как начал работать со своей лошадью. Сначала он научил Ганса считать, держа животное за ногу, показывая ему цифру и постукивая копытом нужное количество раз. Вскоре Ганс уже озвучивал простые суммы. Затем фон Остен ввел доску с написанным алфавитом, и Ганс научился отстукивать цифру для каждой буквы на доске. После двух лет обучения фон Остен был поражен тем, как животное хорошо усваивало интеллектуальные концепции. Поэтому он взял Ганса с собой в дорогу в качестве доказа-

тельства, что животные могут рассуждать, и стал вирусной сенсацией прекрасной эпохи.

Тем не менее, многие люди отнеслись к этому скептически, и совет Германии по образованию создал следственную комиссию для проверки научных утверждений фон Остена. Психолог и философ Карл Штумпф и его помощник Оскар Пфунгст возглавили комиссию по Гансу. Также в нее вошли управляющий цирком, отставной школьный учитель, зоолог, ветеринар и кавалерийский офицер. Однако после продолжительной проверки способностей Ганса, как в присутствии его дрессировщика, так и без него, лошадь сохраняла свой рекорд правильных ответов, и комиссия не смогла найти никаких доказательств обмана. Как позже написал Пфунгст, Ганс выступал перед «тысячами зрителей, любителей лошадей, тренеров первого ранга, и ни один из них в течение многих месяцев наблюдений не смог обнаружить никакого особого «сигнала» между вопрошающим и лошадью. [2]

Комиссия пришла к выводу, что методы, применяемые к обучению Ганса, более схожи с «обучением детей в начальной школе», чем с дрессировкой животных, и «достойны научной экспертизы» [3]. Однако Штумпф и Пфунгст по-прежнему сомневались. В частности, их беспокоил один факт: когда спрашивающий не знал ответа или стоял далеко, Ганс редко отвечал правильно. Пфунгст и Штумпф задумались: а не подавал ли Гансу ответы некий непреднамеренный сигнал?



Вильгельм фон Остен и Умный Ганс

Как описал Пфунгст в своей книге 1911 года, интуиция их не подвела: поза, дыхание и выражение лица человека, задающего вопрос, едва уловимо менялись в тот момент, когда Ганс достигал правильного ответа, тем самым побуждая его остановиться. [4] Позже Пфунгст проверил эту гипотезу на людях и подтвердил догадку. Больше всего в этом открытии его восхитило то, что люди, задающие вопросы, как правило, не знали, что сами давали подсказки лошади. Решения Умного Ганса, писал Пфунгст, основывались на бессознательных сигналах людей [5]. Лошадь была обучена давать те ответы, которые хотел увидеть хозяин, однако зрители не сочли это умение за необыкновенный интеллект.

История Умного Ганса интересна со многих сторон: связь между желанием, иллюзией и действием; развлекательный бизнес; антропоморфизм; возникновение предубеждений и политика интеллекта. Ганс ввел в психологию термин для обозначения особого типа концептуальных ловушек — «Эффект умного Ганса» или «эффект ожидания наблюдателя», с помощью которого описываются влияния непреднамеренных подсказок экспериментаторов на испытуемых. Отношения между Гансом и фон Остеном указывают как на сложные механизмы, посредством которых предубеждения проникают в системы, так и на увязание людей в изучаемых явлениях. В настоящее время история Ганса используется в машинном обучении и служит предостерегающим напоминанием о том, что далеко не всегда можно быть точно уверенным в достоверности полученных моделью данных. [6] Даже система, которая, казалось бы, демонстрирует впечатляющие результаты в процессе обучения, иногда делает ужасные прогнозы на основе совершенно новой информации.

Отсюда вытекает главный вопрос книги: как «создается» интеллект, и какие ловушки могут возникнуть? На первый взгляд, история Умного Ганса — это пример того, как один человек создал интеллект, учив лошадь следовать подсказкам и подражать человеческому разуму. Но с другой стороны мы видим, что практика создания интеллекта значительно шире. Такое начинание требовало подтверждения со стороны множества институтов, включая академические круги, школы, науку, общественность и военных. Более того, фон Остен и его удивительная лошадь обрели рынок — эмоциональные и экономические

инвестиции, которые стимулировали туры, газетные статьи и лекции. Были сформированы бюрократические инстанции, чтобы измерить и проверить и измерить способности лошади. Отсюда следует вывод, что совокупность финансовых, культурных и научных интересов сыграла свою роль в создании интеллекта Ганса, и все кругом были заинтересованы в его уникальности.

Здесь появляются две различные мифологемы. Первая из них заключается в том, что нечеловеческие системы (будь то компьютеры или лошади) являются аналогами человеческого разума. Эта точка зрения предполагает, что при достаточной подготовке или достаточных ресурсах человекоподобный интеллект может быть создан с нуля, без учета фундаментальных межличностных отношений, и помещен в рамки более широкой экологии. Второй миф заключается в том, что интеллект — это нечто, существующее независимо; он является естественным и не сопряжен с социальными, культурными, историческими и политическими силами. На самом же деле концепция интеллекта на протяжении веков наносила огромный вред и использовалась для оправдания отношений господства — от рабства до евгеники. [7]

Эти мифологемы особенно сильны в области искусственного интеллекта, где вера в то, что человеческий интеллект может быть формализован и воспроизведен машинами, с середины двадцатого века стала аксиомой. Подобно тому, как интеллект Ганса считался схожим с человеческим и восприимчивым к обучению, так и системы искусственного интеллекта неоднократно описывались как простые, но человекоподобные формы.

В 1950 году Алан Тьюринг предсказал, что «к концу столетия употребление слов и общее мнение людей изменится настолько, что можно будет говорить о мышлении машин, даже не опасаясь возражений» [8]. Математик Джон фон Нейман в 1958 году утверждал, что функция человеческой нервной системы «на первый взгляд цифровая» [9]. Профессор Массачусетского технологического института Марвин Мински однажды ответил на вопрос о том, могут ли машины думать, сказав: «Конечно, машины могут думать; мы ведь тоже машины, только «мясные»» [10]. Однако нашлись и те, кто не мог согласиться с данной теорией. Джозеф Вейценбаум, ранний изобретатель ИИ и создатель первой программы чат-бота, известной как ELIZA, считал, что

представление о человеке как о простой системе обработки информации является слишком упрощенным понятием интеллекта и порождает «извращенную фантазию», будто ученые ИИ создадут машину, которая сможет обучаться наподобие ребенка» [11].

Это был один из основных споров в истории искусственного интеллекта. В 1961 году в Массачусетском технологическом институте состоялся знаменательный цикл лекций под названием «Управление и компьютер будущего». Звездный состав ученых-компьютерщиков, включая Грейс Хоппер, Дж. К.Р. Ликлайдера, Марвина Мински, Аллена Ньюэлла, Герберта Саймона и Норберта Винера, обсуждал стремительные достижения в области цифровых вычислений. В заключение Джон Маккарти смело заявил, что различия между человеческими и машинными задачами иллюзорны. Просто существуют некоторые сложные человеческие задачи, которые требуют больше времени для формализации и решения машинами [12].

Однако профессор философии Хьюберт Дрейфус выступил с возражением, обеспокоенный тем, что собравшиеся инженеры «даже не рассматривают возможность того, что мозг обрабатывает информацию совершенно иначе, чем компьютер [13]». В своей более поздней работе «Чего не могут вычислительные машины» Дрейфус отметил, что человеческий интеллект и опыт в значительной степени зависят от многих бессознательных и подсознательных процессов, тогда как компьютеры требуют, чтобы все процессы и данные были явными и формализованными [14]. Следовательно, формальные аспекты интеллекта должны быть абстрагированы, устранены или адаптированы для компьютеров, что делает их неспособными обрабатывать информацию так, как это делают люди.

С 1960-х годов в ИИ многое изменилось, включая переход от символических систем к недавней волне шумихи вокруг методов машинного обучения. Во многом споры о способностях ИИ были забыты, а скептицизм сошел на нет. С середины 2000-х годов ИИ быстро развивался как научная область и как индустрия. В настоящее время небольшое число мощных технологических корпораций развертывают системы ИИ в планетарном масштабе, и их системы снова называют сравнимыми или даже превосходящими человеческий интеллект.

Однако история об Умном Гансе напоминает нам о том, насколько узко мы рассматриваем или признаем интеллект. Ганса учили имитировать задачи в очень ограниченном диапазоне: сложение, вычитание и отстукивание слов. Ганс демонстрировал выдающиеся способности в межвидовом общении, публичных выступлениях и значительном терпении, но все это *не было* признано интеллектом. По словам автора и инженера Эллен Ульман, убеждение, будто разум подобен компьютеру и наоборот, «на несколько десятилетий заразило мышление в области компьютерных и когнитивных наук», создав своего рода первородный грех [15]. Это идеология картезианского дуализма в искусственном интеллекте: где ИИ понимается узко, как развоплощенный интеллект, отстраненный от любого отношения к материальному миру.

Что такое искусственный интеллект?

Давайте зададим простой вопрос: «Что такое искусственный интеллект?» Если вы спросите кого-нибудь на улице, он может упомянуть Siri или Apple, облачный сервис Amazon, автомобили Tesla или поисковый алгоритм Google. Если же вы обратитесь к экспертам в области глубокого обучения, вам дадут технический ответ о том, как нейронные сети организовываются в десятки слоев, получают данные типа метки, которым присваиваются пороговые значения, и они могут классифицировать данные таким образом, который пока что не до конца объясним [16]. В 1978 году, обсуждая экспертные системы, профессор Дональд Мичи описал ИИ как совершенствование знаний, где «может быть достигнута надежность и компетентность кодификации, значительно превосходящая самый высокий уровень, которого когда-либо достигал, а возможно, и может достичь, человек-эксперт без посторонней помощи [17]». В одном из самых популярных учебников по этому предмету Стюарт Рассел и Питер Норвиг утверждают, что ИИ — это попытка понять и создать разумные сущности. «Интеллект в основном связан с рациональными действиями, — утверждают они. — В идеале — интеллектуальный агент предпринимает наилучшие возможные действия в той или иной ситуации» [18].

Каждый способ определения искусственного интеллекта выполняет свою задачу, устанавливая рамки того, как его будут понимать, измерять, оценивать и регулировать. Если ИИ определяется потребительскими брендами для корпоративной инфраструктуры, то маркетинг и реклама предопределили горизонт. Если ИИ рассматривается как более надежная или рациональная система по сравнению с человеком-экспертом, то это предполагает, что ему следует доверять принятие решений в области здравоохранения, образования и уголовного правосудия. Когда в центре внимания оказываются конкретные алгоритмические методы, это говорит о том, что важен только постоянный технический прогресс, без учета вычислительных затрат и будущих последствий для планеты.

Напротив, в этой книге я утверждаю, что ИИ не является ни чем-то искусственным, ни интеллектуальным. Скорее, искусственный интеллект — это воплощение и материал, созданный из природных ресурсов, топлива, человеческого труда, инфраструктуры, логистики, истории и классификаций. Системы ИИ не обладают автономностью, рациональностью или способностью распознавать что-либо без длительного, требующего больших вычислительных затрат обучения с использованием больших массивов данных или предопределенных правил и вознаграждений. Искусственный интеллект, каким мы его знаем, полностью зависит от гораздо более широкого набора политических и социальных структур. И из-за капитала, необходимого для масштабного создания ИИ, и способов видения, которые он оптимизирует, системы ИИ в конечном итоге предназначены для обслуживания существующих доминирующих интересов. В этом смысле искусственный интеллект — это реестр власти.

В этой книге мы рассмотрим, как в самом широком смысле создается искусственный интеллект, а также формирующие его экономические, политические, культурные и исторические силы. Как только мы свяжем ИИ со структурами и социальными системами, мы сможем избавиться от представления, будто искусственный интеллект — это исключительно техническая область. На фундаментальном уровне ИИ — это технические и социальные практики, институты и инфраструктуры, политика и культура. Вычислительный разум и вопло-

шенная работа глубоко взаимосвязаны: системы ИИ как отражают, так и производят социальные отношения и понимание мира.

Стоит отметить, что термин «искусственный интеллект» иногда вызывает дискомфорт в сообществе компьютерных наук. Это словосочетание то входит, то выходит из моды на протяжении десятилетий, и используется больше в маркетинге, чем исследователями. В технической литературе чаще используется термин «машинное обучение». Тем не менее, номенклатура ИИ нередко используется в период подачи заявок на финансирование, когда венчурные капиталисты приходят с чековыми книжками, или когда исследователи стремятся привлечь внимание прессы к новому научному результату. Термин ИИ то используется, то снова отвергается, поэтому его значение постоянно меняется. Что касается меня, то я использую ИИ, говоря о массивной индустриальной формации, включающей политику, труд, культуру и капитал. Когда я говорю о машинном обучении, я имею в виду ряд технических подходов (которые, по сути, также являются социальными и инфраструктурными, хотя об этом редко упоминают).

Между тем, существуют значительные причины, по которым данная область была сосредоточена на технических аспектах — алгоритмических прорывах, постепенном совершенствовании продуктов и повышении удобства. Структуры власти на пересечении технологий, капитала и управления хорошо поддаются узкому, абстрактному анализу. Чтобы понять, каким образом ИИ приобретает фундаментально политический характер, нам нужно выйти за рамки нейронных сетей и статистического распознавания образов, и спросить: что оптимизируется, для кого, и кто принимает решения? Затем мы можем проследить последствия этого выбора.

Взгляд на искусственный интеллект как на атлас

Чем же атлас может помочь нам понять принципы создания искусственного интеллекта? Атлас — это необычный тип книги. Он представляет собой собрание разрозненных частей с картами, разрешение которых варьируется от спутникового обзора планеты до подробного

изображения архипелага. Открывая атлас, вы, возможно, ищете конкретную информацию о каком-то месте, а может быть, вы блуждаете, следуя любопытству, и находите неожиданные пути и новые перспективы. Как отмечает историк науки Лоррейн Дастон, все научные атласы стремятся приучить глаз, сфокусировать внимание наблюдателя на конкретных деталях и значимых характеристиках [19]. Атлас представляет определенную точку зрения на мир с отпечатком науки — масштабами, соотношениями, широтами и долготами — и чувством формы и последовательности.

Однако атлас — это в равной степени акт творчества, некое субъективное, политическое и эстетическое вмешательство, сродни научной коллекции. Французский философ Жорж Диди-Юберман считает атлас чем-то, что живет в эстетической парадигме визуального и эпистемической парадигме знания. Задействуя и то, и другое, он подрывает идею о том, что наука и искусство когда-либо полностью разделялись [20]. Вместо этого атлас предлагает нам возможность пересмотреть мир, по-разному связать фрагменты и «снова собрать его воедино, не думая о том, что мы подводим итоги или исчерпываем его». [21]

Мой любимый пример о полезности картографического подхода принадлежит физику и критику технологий Урсуле Франклин: «Карты представляют собой целенаправленную деятельность: они призваны быть полезными, помогать путешественнику и преодолевать разрыв между известным и еще неизвестным; они являются свидетельством коллективного знания и проницательности» [22].

Карты предлагают нам компендиум открытых путей, общих способов познания, которые можно смешивать и комбинировать для создания новых взаимосвязей. Но существуют также карты господства, те национальные карты, на которых территория вырезана вдоль линий разлома власти: от прямого вмешательства при проведении границ через спорные пространства до выявления колониальных путей империй. Ссылаясь на атлас, я хочу сказать, что нам нужны новые способы понимания империй искусственного интеллекта. Нам нужна теория ИИ, учитывающая государства и корпорации, которые управляют им и доминируют над ним; добычу полезных ископаемых, оставляющую отпечаток на планете; массовый сбор данных; а также

глубоко неравные и все более эксплуататорские методы труда, которые его поддерживают. Таковы меняющиеся тектоники власти в ИИ. Топографический подход предлагает различные перспективы и масштабы, выходящие за рамки абстрактных обещаний искусственного интеллекта или новейших моделей машинного обучения. Цель состоит в том, чтобы понять ИИ в более широком контексте, пройдя через множество различных ландшафтов вычислений и увидев, как они связаны между собой. [23]

Атласы актуальны и в другом смысле. Область ИИ явно пытается запечатлеть планету в удобочитаемой для вычислений форме. И это не метафора, а прямое стремление индустрии. Индустрия ИИ создает и нормализует собственные карты, как централизованный взгляд на человеческое движение, общение и труд. Некоторые ученые в области ИИ заявили о желании захватить мир и вытеснить другие формы познания. Профессор ИИ Фей-Фей Ли описывает свой проект *ImageNet* как процесс, направленный на «нанесение на карту всего мира объектов» [24]. В своем учебнике Рассел и Норвиг описывают искусственный интеллект следующим образом: «Механизм, относящийся к любой интеллектуальной задаче; это поистине универсальная область» [25]. Один из основателей искусственного интеллекта и ранний экспериментатор в области распознавания лиц Вуди Бледсоу выразился наиболее прямолинейно: «В долгосрочной перспективе ИИ — это единственная наука» [26]. Идея состоит в том, чтобы не создать атлас мира, а стать атласом. Этот колонизаторский импульс централизует власть в сфере ИИ: он определяет, как измеряется и определяется мир, одновременно отрицая, что это по своей сути политическая деятельность.

Не претендуя на универсальность, книга, которую вы держите в руках, представляет собой частичный отчет. Увлекая вас в мои исследования, я надеюсь показать вам, как формировались мои взгляды. Мы столкнемся с хорошо посещаемыми и менее известными ландшафтами вычислений: шахтами, длинными коридорами энергопоглощающих центров обработки данных, архивами, базами данных изображений и освещенными ангарами. Эти места включены не только для иллюстрации материальной конструкции ИИ и его идеологии, но и для того, чтобы «осветить неизбежно субъективные и политиче-

ские аспекты картирования и предоставить альтернативу гегемонистским и авторитетным подходам», как пишет исследователь медиа Шеннон Мэттерн. [27]

Модели понимания систем уже давно опираются на идеалы прозрачности. Как я писала вместе с исследователем СМИ Майком Анани, способность видеть систему иногда приравнивается к способности знать, как она работает и как ею управлять [28]. Но эта тенденция имеет серьезные ограничения. В случае с ИИ у нас нет «черного ящика», нет секрета, который можно разоблачить, а есть множество переплетенных систем власти. Полная прозрачность является невозможной целью. Скорее, мы лучше понимаем роль ИИ в мире, изучая его материальную архитектуру, контекстную среду и преобладающую политику, а также прослеживая, как они связаны между собой.

Мои размышления опираются на такие дисциплины, как исследование науки и технологий, право и политическая философия, а также на опыт работы в академических кругах и в промышленной исследовательской лаборатории ИИ на протяжении почти десяти лет. За эти годы многие коллеги и сообщества изменили мой взгляд на мир: составление карты — это всегда коллективное занятие, и данная книга не является исключением [29]. Я благодарна ученым, создавшим новые способы понимания социотехнических систем, включая Джеффри Боукера, Бенджамина Браттона, Венди Чун, Лоррейн Дагстон, Питера Галисона, Яна Хакинга, Стюарта Холла, Дональда Маккензи, Ахилла Мбембе, Алондру Нельсон, Сьюзен Ли Стар, Люси Сачман, и многим другим. На создание этой книги повлияли многочисленные беседы и чтение последних работ авторов, изучающих политику технологий, включая Марка Андреевича, Руха Бенджамина, Мередит Бруссард, Симону Браун, Джули Коэн, Сашу Костанза-Чок, Вирджинию Юбэнкс, Тарлетона Гиллеспи, Мар Хикс, Тунг-Хуи Ху, Юк Хуи, Сафию Умоджа Ноубл и Астру Тейлор.

Как и любая книга, моя работа возникла на основе жизненного опыта, что накладывает свои ограничения. Поскольку последние десять лет я жила и работала в США, мое внимание сосредоточено на западной индустрии ИИ. И все же я не ставлю перед собой цель создать полный глобальный атлас: сама эта идея наводит на мысль о захвате и колониальном контроле. Взгляд любого автора основывается