

ГЛАВА 1

Понимание больших языковых моделей и систем

В эпоху до появления смартфонов пользователи мобильных телефонов сталкивались с проблемой: набор текстовых сообщений был *медленным*. Телефоны были ограничены девятиклавишной клавиатурой, а это означало, что для отправки другу сообщения «Can you print out MapQuest directions?»¹ на Nokia 3310 вам нужно было нажать клавишу 84 раза.

В конце 1990-х годов появилось новое решение: T9 (текст на девяти клавишах). Вместо нажатия 2-2-2 для получения буквы С нужно было нажать 2 один раз. Система угадывала, какую из трех букв (А, В или С) вы на самом деле хотите ввести, исходя из слова, которое вы составляли.

Например, чтобы набрать «hello», нужно было нажать 4-3-5-5-6, и T9 определяла, что вы имели в виду H-E-L-L-O, а не G-D-J-J-M или любую другую возможную комбинацию.

Это нововведение ускорило набор для среднестатистического пользователя, но упомянутая система предиктивного ввода текста была придумана не для этого. В конце 1990-х годов американский стартап Tegic Communications начал работать над методами ввода данных для людей с нарушениями моторики, которым было трудно пользоваться традиционными клавиатурами. Инженеры компании искали, как создать эффективный ввод текста, используя всего восемь отслеживаемых положений глаз, что значительно снизило физические усилия, необходимые для составления слов. Tegic обнаружила, что тот же принцип поможет усовершенствовать обычные мобильные телефоны, и перепрофилировала свою работу по обеспечению доступности на удовлетворение растущего спроса на эффективную мобильную электронную почту и текстовые сообщения, что привело к созданию одной из ведущих мобильных систем ввода данных начала 2000-х годов.

¹ «Можешь распечатать маршрут MapQuest?»

Ранние предиктивные языковые модели

В некотором смысле T9 была похожа на большую языковую модель (large language model, LLM): это был инструмент для обоснованных (на основе входных данных) предположений о намерениях человека, выраженных в письменной речи. Система поддерживала словарь распространенных слов, ранжированных по частоте, что сужало круг возможных вариантов по мере ввода текста. Когда несколько слов имели одинаковую последовательность клавиш, например «good» и «home» имели последовательность 4-6-6-3, T9 сначала предлагала наиболее часто используемое в естественном языке слово.

Однако по сравнению с другими технологиями T9 была относительно небольшой: вся система, включая словарь, имела размер файла меньше, чем размер одной фотографии. Тем не менее в этих рамках она обеспечивала впечатляющий и важный пользовательский опыт. Те же инженеры впоследствии создали Swype (рис. 1.1), клавиатуру с управлением жестами для сенсорных телефонов, инструмент, позволявший пользователям составлять слова, плавно проводя пальцем от буквы к букве, и оказавший влияние почти на все последующие клавиатуры.



Рис. 1.1. Swype, ранняя версия интерфейса предиктивного ввода слов для распознавания жестов при наборе текста

Таким образом, хотя LLM-системы по своему подходу напоминают ранние текстовые предсказатели, на практике они представляют собой совершенно иное явление. Если T9 ускоряла набор текста с помощью эффективных

предположений, то LLM-системы, как вы увидите в этой книге, способны превратить текстовый запрос в эссе, резюме, код и многое другое. Тем не менее основная цель остается неизменной: эти предиктивные системы представляют собой методы прогнозирования последовательности слов на основе паттернов использования, выявленных в результате обработки огромных объемов текстовых данных, написанных людьми. LLM-системы кажутся каким-то волшебством, но на самом деле они лишь отражают закономерности и логику, заложенные в человеческом языке, а не более глубокое понимание мира вокруг.

Этот подход полезен, поскольку привязывает наши проектные решения к потребности человека в общении, не позволяя ему отвлекаться на технические возможности или обещания о предполагаемом появлении искусственного общего интеллекта (artificial general intelligence, AGI). Если мы упускаем из виду этот человекоцентричный принцип, мы создаем функции ИИ, которые могут демонстрировать впечатляющие технические достижения, но не решают реальные проблемы коммуникации, с которыми сталкиваются люди, — распространенная жалоба, которую мы слышим от пользователей продуктов ИИ.

Чтобы понять, как мы пришли к современным системам ИИ, почему они ставят такие уникальные задачи разработчикам и почему существуют некие предположения, которые привносят пользователи, нам необходимо проследить эволюцию ИИ. Путь от ранних вычислительных экспериментов до современных языковых моделей показывает фундаментальные сдвиги в нашем понимании машинного интеллекта. Каждая эпоха создавала свои собственные предположения о том, что могут делать компьютеры и как люди должны с ними взаимодействовать, и эти предположения до сих пор влияют на проектные решения, принимаемые сегодня.

Краткая история ИИ

Еще 10 лет назад об искусственном интеллекте почти ничего не было слышно. Я пришла в эту область благодаря работе над проектом TensorFlow, одной из первых моделей Google с открытым исходным кодом. В рамках одного из проектов Google попросила мою команду создать интерактивные приложения, которые помогли бы пользователям понять базовые технологии трех новейших моделей. Первая, хорошо известная PoseNet, могла определять позу человека по изображениям или видео и преобразовывать их в данные для отслеживания положения тела. Вторая могла анализировать изображение человеческого лица и оценивать эмоции — радость, грусть, гнев и т. д. Третья представляла собой модель обработки естественного языка

(natural language processing, NLP). Чтобы объяснить, как работают модели обработки естественного языка, мы разработали игру (рис. 1.2), в которой пользователь мог перемещаться по галактике, вводя слово, находящееся в наибольшей семантической близости к другому слову. Например, если пользователь хотел отправиться на планету А и соответствующее слово для планеты А было «биология», он мог ввести «форма жизни», что обеспечивало 90%-ную связь между двумя словами и задавало траекторию движения корабля в правильном направлении.

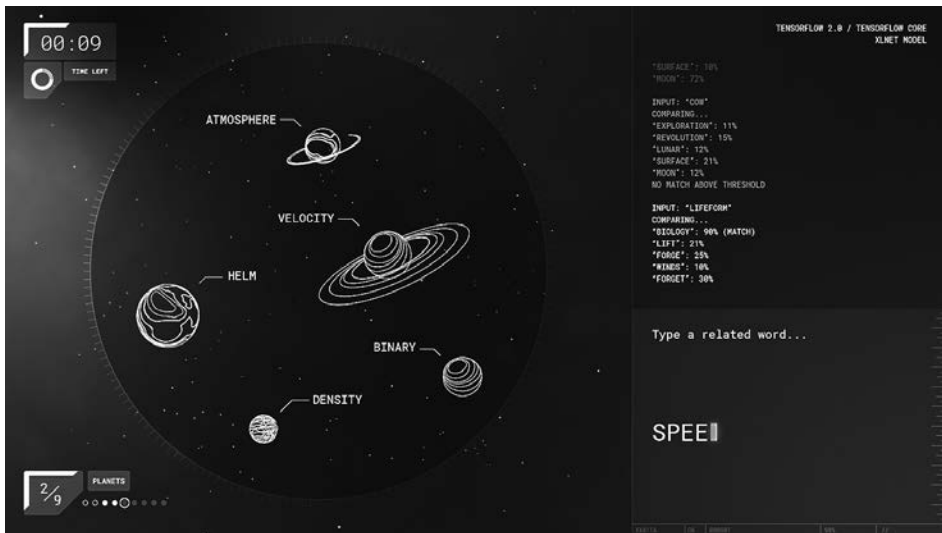


Рис. 1.2. Startographer, интерактивная игра, в которой пользователи прокладывают курс в космическом пространстве, приблизительно определяя расстояние между двумя словами

Наша игра оказалась удачной, потому что помогала пользователям концептуализировать слова, существующие в непосредственной близости друг от друга в многомерном пространстве. Но ее разработка представляла собой сложную задачу. Из всех проектов модель обработки естественного языка оказалась для нас самой сложной в плане организации пользовательского опыта. Мы создали ее последней, потому что поняли, что нам потребуется больше времени на понимание задачи, чем на реализацию самого решения. Я помню, как проводила дни с инженерами проекта, экспериментируя с различными вариантами, каждый раз пытаясь лучше понять, *на что* способен наш продукт.

Эта фаза экспериментирования — изучение возможностей и их преобразование в интерфейс — до сих пор остается одной из самых сложных при разработке продуктов на основе искусственного интеллекта. Для того чтобы

помочь пользователям создать точные ментальные модели системы, необходимо хорошо понимать базовую технологию, а также уделять приоритетное внимание пользовательскому опыту, соотнося его с потребностями бизнеса и учитывая возможные глубокие изменения в будущем — все это мы рассмотрим в главах книги.

И хотя современные модели значительно превосходят ту модель NLP, с которой я работала в 2018 году, это не решает проблему предвзятых и часто неточных представлений пользователей о продукте. Действительно, эта трудность существовала за десятилетия до появления современных языковых моделей, возникая всякий раз, когда появлялись какие-нибудь новые «интеллектуальные» возможности. Это один из важнейших уроков работы с ИИ, который объясняет, почему инженеры сталкиваются с проблемой, выходящей за рамки типичного проектирования программного обеспечения: пользователи обращаются к продуктам ИИ, уже имея необоснованные предположения, которые зачастую совершенно не соответствуют реальному положению дел, и часто людям не хватает поддержки, чтобы открыть для себя *реальные* возможности продукта.

Пользователи руководствуются предвзятыми представлениями об ИИ, сформированными научной фантастикой, маркетинговой шумихой, возможно, даже собственным опытом работы с несвязанными, но похожими системами, которые могли работать совершенно иначе (например, поиск Google). Средства массовой информации постоянно распространяют истории об абсолютном общем интеллекте, способном уничтожить планету, — но это не имеет ничего общего с технологиями, с которыми люди сталкиваются на практике. Понимание пути от теоретического к реальному ИИ полезно в контексте пользовательского опыта использования продукта, поскольку тернистый путь прогресса всегда сопровождают легенды, которые только вызывают всеобщую путаницу.

Начало: символы

С появлением программируемых цифровых компьютеров в 1940-х годах возникло множество предположений о том, могут ли они на самом деле *мыслить*. Математик и ученый Алан Тьюринг попытался дать точный ответ на этот вопрос в своем эссе 1950 года «Вычислительные машины и разум»¹. Он предложил то, что назвал *игрой в имитацию*, теперь широко известной как тест Тьюринга (рис. 1.3). Человек-судья (спрашивающий) обменивался письменными сообщениями с двумя невидимыми участниками, одним человеком и одной машиной. Если судья не мог надежно отличить их друг от друга, считалось, что машина обладает интеллектом.

¹ https://ru.wikipedia.org/wiki/Вычислительные_машины_и_разум

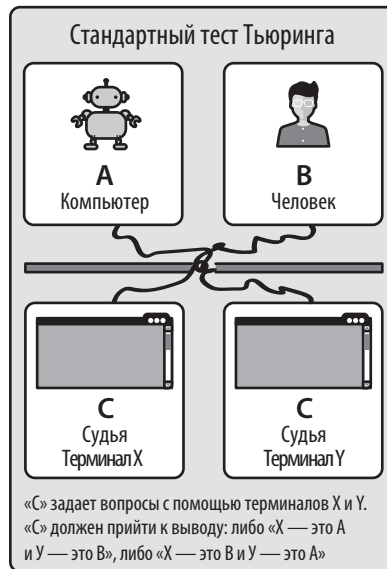


Рис. 1.3. Иллюстрация оригинального теста Тьюринга, в котором человек-судья пытается отличить компьютер от человека, основываясь исключительно на их текстовых ответах, при этом цель компьютера состоит в том, чтобы успешно выдать себя за человека (взято из статьи С. Г. Стерретта (S.G. Sterrett) «The Genius of the “Original Imitation Game”»)

Эта идея оказалась не вполне состоятельной. Теоретически оценивая все виды интеллекта, тест Тьюринга на самом деле ставил вопрос очень узко: может ли машина убедительно имитировать человека в текстовом диалоге? В статье не учитывалось, насколько сильно пользователи будут *переносить* интеллект в разговор, звучащий как человеческий, и насколько сильно отличаются и далеки друг от друга *реальный* интеллект и разговор, похожий на человеческий. Технические возможности, воспроизводящие человеческую речь, оказались совершенно иными, чем предполагалось изначально, по сравнению со способностями к независимому рассуждению.

Компьютерная программа ELIZA, разработанная в Массачусетском технологическом институте в середине 1960-х годов Джозефом Вайценбаумом (Joseph Weizenbaum), была одной из первых попыток исследовать это противоречие. Вайценбаум создал ее как эксперимент в области обработки естественного языка, показав, как компьютер может вести диалог, следуя простым правилам распознавания ключевых слов и генерации ответов. Сценарий «DOCTOR» был разработан для имитации работы психотерапевта и воспроизводил слова пользователя с небольшими преобразованиями. На рис. 1.4 можно прочитать типичный диалог с ELIZA.

```

Welcome to
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II    ZZ    AA   AA
EEEEEE LL      II    ZZZ   AAAAAA
FF      LL      II    ZZ    AA   AA
EEEEEE LLLLLL IIII  ZZZZZZ  AA   AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:

```

Рис. 1.4. Типичный разговор пользователя с программой ELIZA¹

Чтобы создать иллюзию разговора, ELIZA сканировала ввод пользователя на наличие ключевых слов, применяла правила сопоставления шаблонов и собирала фрагменты текста в шаблонные ответы. Например, если она обнаруживала слово «бойфренд», она могла повторить этот термин в вопросе, требующем размышления. Если ключевое слово не распознавалось,

¹ Текст на скриншоте:

ELIZA — имитация психотерапевта в духе подхода Карла Роджерса.

Оригинальная программа была описана Джозефом Вейценбаумом в 1966 году.

Эта реализация — Норберта Ландштайнера, 2005 год.

ELIZA: Вас что-то тревожит?

ЧЕЛОВЕК: Все мужчины одинаковы.

ELIZA: Как вы думаете, с чем это связано?

ЧЕЛОВЕК: Вечно пристаю и все такое.

ELIZA: Можете привести конкретный пример?

ЧЕЛОВЕК: Я здесь из-за своего бойфренда.

ELIZA: Для вас важно, что вы пришли сюда из-за своего бойфренда?

ЧЕЛОВЕК: Он говорит, что я почти постоянно нахожусь в депрессии.

ELIZA: Мне жаль слышать, что у вас депрессия.

ЧЕЛОВЕК: Это правда. Я несчастна.

ELIZA: Можете объяснить, отчего вы несчастны?

ЧЕЛОВЕК: ...

она по умолчанию отвечала общими вопросами, такими как «Расскажите подробнее».

Возможность отвечать заставила пользователей быстро наделить инструмент чертами человека: они начали сопереживать программе, приписывая ей понимание, знания и сочувствие, тогда как она лишь оперировала фрагментами их собственных высказываний. Сам Вайценбаум был обеспокоен тем, что люди делают выводы о наличии когнитивных способностей там, где их нет. В своих записях он описал один тревожный случай взаимодействия его секретарши с ELIZA:

Моя секретарша долгое время наблюдала за моей работой над этой программой. Однажды она попросила разрешения пообщаться с системой. Конечно, она понимала, что взаимодействует с машиной. И все же, напечатав несколько предложений, она повернулась ко мне и сказала: «Не могли бы вы выйти из комнаты, пожалуйста?»

Это явление позже получило название *эффекта ELIZA*: тенденция проецировать свойственные человеку качества на лингвистические результаты, которые лишь имитируют понимание. Это логично — на протяжении большей части истории человечества, за редкими исключениями (лирохвосты¹ и т. д.), говорить могли только люди. Поскольку язык является исключительно человеческим способом общения, предполагается, что он подготавливает мозг к освоению других человеческих умений — важное обстоятельство для тех, кто создает инструменты, имитирующие человеческий язык.

Проекты ELIZA и ему подобные возникли из общего желания выявить определенный набор правил, которые сделали бы сгенерированный язык убедительно реалистичным. В 1950-х годах Ноам Хомский (Noam Chomsky) предположил, что люди обладают врожденной системой синтаксических правил, которые управляют языком и смыслом. Его пример — «Бесцветные зеленые идеи яростно спят» — иллюстрирует парадокс: предложение грамматически правильно, но настолько очевидно бессмысленно, что ни один взрослый человек никогда бы его не произнес. Лингвистическая теория Хомского и эксперимент ELIZA определили эпоху осознания того, что язык включает в себя как структуру, так и интерпретацию, заставив исследователей перейти от планирования на основе правил к изучению статистических методов, которые могли бы более точно моделировать фактическое использование языка.

¹ Птица лирохвост обладает уникальной способностью к звукоподражанию и умело подражает голосам других птиц и крикам млекопитающих.

Но исследователи ИИ застряли в попытках жестко запрограммировать логику машины в надежде на создание истинного интеллекта. Эта неспособность выйти за рамки систем, основанных на правилах, привела к первой «зиме ИИ» 1970-х годов. Финансирование иссякло, когда стало понятно, что логика исследователей не работает: они не могли изложить всю бесконечную сложность человеческого языка в алгоритме.

Даже спустя десятилетия чат-боты оставались в тени ELIZA. SmarterChild (рис. 1.5), запущенный в 2001 году и используемый миллионами пользователей AOL Instant Messenger, по сути, был ELIZA с большим словарем. Он мог сообщать расписание киносеансов и цены на акции, но его разговорные возможности принципиально не улучшились.

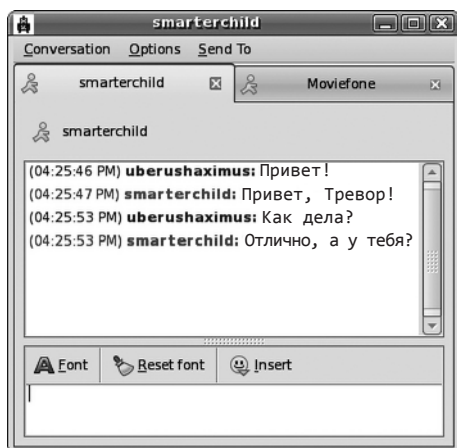


Рис. 1.5. SmarterChild от AOL, ранний чат-бот¹

Для разработки модели, которая могла бы более разумно синтезировать человеческий язык и поведение, были изобретены новые статистические подходы. В то время как более старые методы использовали простой подсчет (как часто артикль «the» встречается перед словом «cat»? Как часто слово «chat» во французском языке соответствует слову «cat» в английском?), эти исследователи пытались изучить более глубокие закономерности с помощью слоев искусственных нейронов. Надежда возлагалась на то, что сеть сможет обнаруживать абстрактные признаки и связи между словами, используя данные для понимания того, что слова «dog» («собака») и «puppu» («щенок») связаны по смыслу, без явного указания на это.

¹ На самом деле SmarterChild не поддерживал русский язык, перевод приведен для удобства читателя.

Однако нейронные сети были вычислительно затратными: обучение даже небольшой сети занимало дни. На оборудовании 1980-х годов простые методы подсчета работали лучше.

Три обстоятельства способствовали появлению современного ИИ: стремительный рост объемов данных, появление быстрых графических процессоров и архитектура трансформеров.

Прежде всего, взрывной рост интернета предоставил программистам огромные массивы так необходимых им обучающих данных. В 1980-х годах считалось удачей, если исследователям удавалось найти несколько тысяч цифровых документов. К 2010 году объемы данных стали колоссальными. Только Википедия содержала 3,5 миллиона статей на английском языке. Google Books отсканировал 20 миллионов текстов. Каждый день пользователи сети X (ранее Twitter) публиковали 500 миллионов твитов. Facebook генерировал миллиарды обновлений статуса. Проект Common Crawl, запущенный в 2011 году, заархивировал весь интернет, создав петабайты необработанного текста, доступного для бесплатного скачивания любому желающему.

Тип создаваемых данных также значительно отличался. В то время как предыдущие датасеты были чистыми и академическими, интернет представлял исследователям сложный человеческий язык: споры на форумах, текстовые сообщения, обзоры продуктов, журналы чатов. Нейронные сети, обученные на этих данных, учились тому, как люди *на самом деле* общаются, а не как пишутся научные статьи.

Следующее нововведение касалось проблемы вычислительной мощности. В 2012 году аспирант Алекс Крижевский (Alex Krizhevsky) принял участие в конкурсе по классификации изображений. В то время как другие участники использовали центральные процессоры, Крижевский использовал две игровые видеокарты NVIDIA GTX 580, созданные для рендеринга таких игр, как *Skyrim*. Его нейронная сеть, названная AlexNet, разгромила конкурентов. В итоге видеокарты, созданные для игр, случайно произвели революцию в мире ИИ. Оказалось, что сценарии использования аналогичны: подобно тому как играм необходимо одновременно определять освещение и физику для объектов, вычисляя миллионы пикселей, нейронным сетям необходимо одновременно вычислять значения миллионов нейронов. Один графический процессор (GPU) содержал тысячи крошечных процессоров, способных работать параллельно, в то время как центральный процессор (CPU) имел, например, восемь мощных процессоров, работающих последовательно. Для нейронных сетей, которые в основном выполняли простые вычисления на огромных массивах данных, GPU были в 50–100 раз быстрее, чем CPU. К 2015 году NVIDIA начала производить графические процессоры специально для ИИ.

Третье и последнее нововведение решило проблему *обработки* такого огромного объема данных: изобретение архитектуры трансформеров, которая решила проблему чтения текста нейронными сетями. До 2017 года сети обрабатывали слова по одному, как при чтении бегущей строки. В 2017 году команда Google опубликовала статью под названием «Attention Is All You Need»¹, в которой была представлена вычислительная концепция трансформеров. Вместо последовательной обработки слов трансформеры рассматривали предложения целиком. Ключевым нововведением стало *внимание* — механизм, позволяющий каждому слову напрямую связываться с каждым другим словом, вычисляя, какие связи наиболее важны. Обработка текста «The cat was happy» («Кот был счастлив») больше не означала чтение слева направо. Глядя на слово «happy», трансформер мог мгновенно проверить: что с ним связано? Он обнаруживал, что «cat» имеет сильную связь, «was» — среднюю, а артикль «the» — почти никакой. Такая параллельная обработка ускорила обучение и позволила сетям понимать гораздо более длинные тексты.

В течение нескольких месяцев все крупные лаборатории ИИ перешли на трансформеры и стали использовать все три инновации для быстрого создания и обучения совершенно нового типа моделей. Google создал BERT (2018), который побил все рекорды в понимании языка. Компания OpenAI разработала GPT (2018) с первоначальными 117 миллионами параметров — регулируемыми значениями внутри нейронной сети, а затем GPT-2 (2019) с 1,5 миллиарда параметров и GPT-3 (2020) со 175 миллиардами параметров. Каждая версия просто масштабировала ту же архитектуру трансформера с большим объемом данных и большей мощностью графического процессора. Сочетание этих факторов оказалось прорывным: трансформеры смогли использовать миллиарды веб-документов, GPU смогли обучать огромные сети, и внезапно машины стали писать эссе, отвечать на вопросы и генерировать код на уровне, близком к человеческому.

Однако этот прорыв в разработке моделей и повышении их компетентности вызвал определенные опасения, и исследователи констатировали очевидное несоответствие между ожиданиями общественности от ИИ и тем, что на самом деле происходило «под капотом». В своей статье 2021 года «On the Dangers of Stochastic Parrots» (<https://oreil.ly/IvNUg>) исследователи Эмили Бендер (Emily Bender), Анджелина Макмиллан-Мейджор (Angelina McMillan-Major), Тимнит Гебру (Timnit Gebru) и Маргарет Митчелл (Margaret Mitchell) высказали опасение, что гонка за созданием все более крупных моделей рискует упустить из виду важнейшие проблемы: эти

¹ «Внимание — это все, что вам нужно» (https://ru.wikipedia.org/wiki/Внимание_—_это_все,_что_вам_нужно)

системы могут усиливать предвзятость, или смещение (bias), присутствующую в их обучающих данных, потреблять огромное количество энергии с экологическими последствиями и создавать иллюзию понимания, просто комбинируя фрагменты из своего обучающего корпуса. Их работа показала, как тот же масштаб, который обеспечил впечатляющие возможности, также сконцентрировал власть в руках компаний, обладающих ресурсами для обучения масштабных моделей, одновременно затрудняя понимание или исправление проблемного поведения, заложенного глубоко в миллиардах параметров.

К 2026 году ландшафт ИИ определяется сложной и быстро развивающейся экосистемой, выходящей далеко за рамки ранних генеративных моделей. Рынок является высококонкурентным, и на нем представлены такие крупные игроки, как GPT-4 от OpenAI, продемонстрировавший впечатляющие результаты в профессиональных задачах и программировании; модели Claude от Anthropic, которые ставят во главу угла безопасность за счет конституционального обучения ИИ (двухэтапный процесс, использующий набор принципов, называемых «конституцией», вместо обратной связи от человека для обучения моделей ИИ); и модели Gemini от Google, которые развивают мультимодальные возможности, бесшовно интегрирующие текст, изображения и код. Однако центральной проблемой в эту эпоху остается значительный разрыв между технической реальностью функционирования систем и их пониманием обществом. Дизайнеры должны постоянно работать над преодолением устойчивой тенденции, при которой пользователи проецируют интеллект человеческого уровня на системы, являющиеся по своей сути статистическими, и решать такие проблемы, как непредсказуемые сбои и «галлюцинации», чтобы укрепить доверие и сформировать понимание истинных возможностей ИИ.

Пользовательский опыт в эпоху ИИ

Что означает, когда мы говорим, что пользователи привносят предположения в продукты ИИ? В качестве отправной точки важно четко понимать *типы* недоразумений, которые возникают у пользователей.

Как упоминалось ранее, часто можно увидеть, как пользователи проецируют на модель человекоподобное понимание, полагая, что ИИ улавливает контекст, эмоции и нюансы таким образом, который выходит за рамки его возможностей. Это часто сочетается с предположением о фактичности, что заставляет пользователей считать сгенерированную информацию правильной и надежной, не учитывая склонность модели к галлюцинациям или генерации правдоподобных ошибок. Еще больше усложняет ситуацию

предположение о согласованности и надежности. Ожидание, что один и тот же промпт всегда будет давать один и тот же правильный результат, игнорирует вероятностную природу ИИ, которая приводит к изменчивости результатов.

К другим распространенным заблуждениям относится предположение о том, что модель обладает всей полнотой информации: пользователи считают, что ИИ имеет доступ ко всем сведениям и текущим событиям независимо от порогового значения обучающих данных; а также предположение о непредвзятости результатов, которое не учитывает тот факт, что обучающие данные модели неизбежно отражают и могут воспроизводить человеческие предубеждения.

Фундаментальный конфликт между априорными предположениями и имеющимися возможностями ИИ может помешать пользователям открыть для себя *реальные* функции ИИ, которые могли бы оказаться наиболее полезными. Когда пользователи предполагают у ИИ наличие понимания, подобного человеческому, они начинают неправильно применять искусственный интеллект к сложным задачам рассуждения, что приводит к разочарованию и недооценке ИИ как инструмента для поиска информации, обобщения и составления документов. Аналогично вера в присущую модели фактичность заставляет пользователей пропускать важные шаги, такие как проверка фактов, в результате чего ИИ используется в качестве окончательного авторитета, а не умелого партнера для генерации идей. В итоге чрезмерная убежденность в том, что ИИ обладает совершенным знанием, мешает пользователям применять нейросети наиболее эффективным образом: как помощника исследователя, требующего партнерства с человеческим опытом и критическим суждением.

Конфликт между предположениями пользователей и реальными возможностями ИИ значительно усложняется и усугубляется огромными финансовыми инвестициями, которые вкладываются в сектор ИИ. Мировые расходы на ИИ в 2025 году составили 1,5 триллиона долларов (https://oreil.ly/rk_vt) и, как ожидается, будут только расти в 2026 году. Предприятия находятся под огромным давлением, стремясь быстро внедрить функции ИИ, чтобы оправдать и окупить эти огромные капиталовложения, и в результате приоритет отдается скорости, стоимости и немедленному влиянию на рынок, что может усугубить проблемы, возникающие из-за априорных предположений пользователей.

Например, стремление продемонстрировать инновации часто приводит к продвижению функций, которые пользователю не нужны или неинтересны, когда отдается предпочтение технологическим возможностям в ущерб пользе. В некоторых случаях эта проблема усугубляется отсутствием

ориентации на аудиторию: если продукты пытаются удовлетворить потребности слишком многих сценариев использования или целевых групп, не сужая фокус до основной аудитории, то создаваемые функции ИИ могут напрямую противоречить тому, как человек хочет применять инструмент или взаимодействовать с ним. Это несоответствие еще больше запутывает пользователя, затрудняя формирование точной ментальной модели истинных возможностей системы или того, как она может эффективно выполнить задачу.

Помимо давления со стороны бизнеса важную роль в этом конфликте играет широкая среда разработки и внедрения ИИ. Как упоминалось ранее, непрерывный цикл ажиотажа в СМИ, особенно вокруг прорывов в генеративном ИИ, создает представление о его почти магических, безграничных возможностях, ускоряя культурное принятие ИИ как оракула и затрудняя обучение пользователей его гораздо более скучной статистической природе. Это сопровождается отсутствием универсальных стандартов и регулирования. Нарастающий регуляторный вакуум во многих юрисдикциях означает отсутствие обязательных стандартов прозрачности (хотя, как мы увидим, первые законодательные акты уже начали приниматься).

Наконец, сам дизайн интерфейса и пользовательского опыта многих инструментов ИИ усугубляет проблему: чистые, минималистичные и разговорные интерфейсы в некоторых случаях прячут лежащую в их основе сложность и вероятностную природу, укрепляя иллюзию, будто пользователь взаимодействует с разумным существом, и скрывая реальность того, что это сложный статистический механизм.

Для дизайнеров работа по преодолению этого разрыва начинается с понимания своего пользователя и особенностей использования продукта. Это их привычная сфера, но в эпоху ИИ такое понимание выходит за рамки документирования потребностей или сбора запросов на новые функции. Теперь необходимо выяснять, когда пользователи могут неправильно истолковать информацию, какие сигналы могут сбить их с толку и насколько стоит раскрывать детали базового процесса. Иногда ясное объяснение помогает, но столь же часто слишком большое количество подробностей перегружает или отвлекает. Мастерство заключается в том, чтобы определить, где прозрачность укрепляет доверие, а где только создает путаницу, и затем последовательно принимать такие решения, чтобы продукт оставлял ощущение целостной системы.

Здесь также играют важную роль технические навыки и умение оценивать. Дизайнерам не нужно становиться исследователями в области машинного обучения, но им необходима достаточная подготовка, чтобы понимать, что происходит «под капотом». Без этих знаний трудно объяснить поведение

модели, предвидеть возможные сбои в системе или установить четкие ожидания как для пользователей, так и для коллег. Эти навыки позволяют дизайнерам оценивать, какие детали системы следует сглаживать, какие отображать и как представить их таким образом, чтобы это помогало пользователям, а не отвлекало их.

Поскольку продукты на основе ИИ создаются на стыке различных дисциплин, коммуникация внутри организации имеет не меньшее значение. Дизайнерам, инженерам и менеджерам продукта необходим общий язык для описания того, что система может делать, чего она не может и что следует или не следует показывать пользователям. Когда эти стандарты установлены и соблюдаются, пользователи видят согласованную логику во всем продукте. Организации также выигрывают от четкого владения знаниями и практики обмена ими: в быстро развивающейся отрасли продукты на основе ИИ требуют стратегических целенаправленных методов коммуникации и адаптации к технологическим изменениям.

В этом смысле три требования — знание своего пользователя, развитие технических навыков и умения оценивать, а также совершенствование коммуникационных практик — направлены на одну цель: уменьшение двусмысленности и предоставление пользователям возможности понять, что перед ними. Именно эта ясность делает продукты на основе ИИ пригодными для использования, а без нее даже самая продвинутая система рискует быть неправильно понятой или вызвать недоверие. Давайте рассмотрим каждое требование немного подробнее.

Знай своего пользователя

Если интерфейс может быть *любым*, как его проектировать? Много лет назад, работая над проектом для Google Search, я столкнулась с непростой задачей: необходимо было визуализировать результаты поиска. Проблема заключалась в следующем: чтобы сделать категории Google более доступными, дизайнеру нужно было предвидеть результат любого запроса пользователя, отчасти для того, чтобы помочь ему найти различные типы результатов (изображения, авиабилеты, новости и т. д.). Но учитывая, что искать можно *что угодно*, как разработать дизайн, который улучшит конкретный пользовательский опыт? К счастью, Google поделился опубликованными исследованиями, в которых описаны восемь наиболее распространенных типов поисковых запросов (рис. 1.6). Их анализ позволил выявить *типы* задач, которые пользователи обычно выполняют с помощью запросов, и это помогло мне установить более явную связь между потребностями пользователя и возможностями, которые должен предоставлять интерфейс.