

# ВВЕДЕНИЕ

Даже если вы далеки от технологических трендов, все равно у вас мало шансов ни разу не услышать о нейросетях и прорывах в области искусственного интеллекта (ИИ).

Нейросети то, нейросети се... Нейросети массово лишают работы художников. Нейросети обучают управлению самолетом. Нейросети внедряют в чат-боты, чтобы вживую общаться с клиентами. Этим новостей сегодня так много и число их растет настолько быстро, что даже самый ленивый человек хотя бы мельком, но обратит внимание.

А уж тот, кто не ленив и готов увидеть перспективы применения ИИ в бизнесе, работе, образовании, развлечениях и многом другом, и вообще смотрит в оба глаза.

В этой книге мы не станем замахиваться на понятие «нейросети вообще», давать сложные и долгие объяснения или делать другие заумные вещи, которые 95 % читателей просто не поймут и начнут зевать.

**Наша цель куда проще:** мы возьмем один вид нейромоделей (генерация контента) и постараемся подружить вас с ним, перевести ваше общение с ИИ из формата «Что ТЫ такое?» в формат «Привет, поработаем?». Благо у нас (авторов данной книги) для этого есть и наработки, и кое-какой опыт, и собственный инструментарий. Приступим.

## LLM И GPT КАК ЧАСТЬ ИИ

ChatGPT — лишь часть глобального явления под названием «искусственный интеллект». Думаем, будет правильно, если мы опишем структуру этого явления, пусть даже самую суть, — чтобы вы могли понять его.

Если же вам нужно еще больше информации, то вы всегда можете спросить ИИ или найти в поисковиках. Это очень просто. Итак, начнем.

*Искусственный интеллект* (ИИ; от англ. artificial intelligence, AI) — компьютерная технология, которая помогает выполнять задачи, требующие творческого подхода, логического мышления, аналитических способностей и прочих признаков, которыми обладает только человек.

ИИ отличается от обычных программ тем, что может обучаться на основе опыта, менять выводы в зависимости от контекста, анализировать сложные вводные и многое другое.

Архитектура ИИ отчасти похожа на строение человеческого мозга с его нейронными связями. Разница только в том, что в ИИ эти связи — искусственные. Чем сильнее развиваются технологии, чем больше опыта и возможностей накапливается у ИИ и его разработчиков, тем «умнее» становится ИИ, тем больше улучшается качество искусственных связей.

*Нейросети* — нечто вроде подкласса ИИ. Скажем, если ИИ вообще — это семейство кошачьих, то нейросети — это и гепарды, и львы, и пумы, и домашние рыжие Васьки.

Нейросети опираются на общие принципы работы ИИ, но при этом могут кардинально различаться по уровню задач. Есть разные подвиды нейросетей наподобие рекуррентных, многослойных и т. п., но в данной книге мы не будем о них говорить.

Мы будем рассматривать нейросети по уровню задач. Например, нейросети:

- для распознавания голоса;
- для создания изображений;
- для работы с информацией (контентом) и т. п.

А вот уже внутри одного направления нейросетей (например, контентных) и находится то, с чем конкретно мы будем работать, — LLM.

*LLM* (от англ. Large Language Model, большие языковые модели) — это обученные модели для выполнения конкретных задач. LLM можно сравнить с ребенком: чем больше с ним общаются, чем больше фактов он узнает, тем больше учится и развивается, пытается анализировать, находить закономерности, работать с вводными данными и делать выводы.

Разные LLM можно обучать совершенно по-разному. Одни из них огромны и тратят колоссальное количество аппаратных ресурсов, для работы им нужны большие залы, наполненные самыми мощными и современными серверами. Другим же LLM хватит и самых простых «бытовых» мощностей.

Одни LLM узкоспециальны и обучаются под конкретную задачу: например, сводить оборот на основе документов одной компании. Другие готовы обслуживать население целых стран и выполнять тысячи разных задач.

Например, YandexGPT, Google Bard и версии GPT от OpenAI — «родственники». Это модели для работы с информацией (контент, написание кода, анализ текстов и т. п.). Но чем больше мастерства, сил, средств, наконец, вложено в развитие конкретной LLM, тем радикальнее могут различаться результаты: по уровню знаний, внутренней архитектуре, объему контекста и т. д. При этом цели и задачи однотипных LLM примерно похожи.

А вот если вы хотите, например, рисовать, то нужны LLM наподобие Stable Diffusion, «Кандинский» или Midjourney. Для обработки видео вам понадобятся другие модели. И т. д.

Теперь, когда мы немного разобрались, поговорим о линейке конкретных LLM-моделей — GPT (ChatGPT).

*GPT* — передовая линейка генеративных нейромоделей для работы с информацией, которую разрабатывает компания OpenAI. Здесь есть небольшой нюанс: известный многим термин *ChatGPT* (если уж совсем дотошно) — не совсем верное название для всей серии LLM. Так называется лишь знакомый нам интерфейс для работы, похожий на обычный чат.

А вот GPT-3, 4, 5... — это номера конкретных версий моделей компании, где каждая следующая опережает предыдущую по целому ряду показателей.

В этой книге мы будем оперировать словами *ChatGPT*, «ИИ», «нейросеть», «модель» и т. п., имея в виду в первую очередь GPT от OpenAI. Просто потому, что на момент написания книги это явный лидер, который задает тон.

Но вообще мы создаем книгу универсального промтинга, не привязываясь к конкретному номеру модели и даже к конкретной LLM (что такое промтинг, мы расскажем чуть позже). Самый смак в том, что эти правила одинаковы для всех аналогичных LLM, поэтому подойдут и для других моделей. Почему — об этом еще поговорим, а пока продолжим.

Линейка моделей GPT сегодня опережает другие похожие решения. В OpenAI и начали раньше, и денег, видимо, хватало, и люди там талантливые. Словом, лидеры.

ChatGPT<sup>1</sup> постоянно развивается, всегда на несколько шагов впереди. Пока другие LLM доходят до его вчерашнего уровня, он уже машет рукой из завтрашнего дня.

---

<sup>1</sup> Далее для единообразия мы будем обращаться к ИИ (GPT, ChatGPT) в мужском роде.

Общение человека с ChatGPT ведется с помощью текстовых запросов (prompts), которые мы далее будем называть *промтами* или *подсказками*.

☞ В тексте они обозначаются вот таким шрифтом.

Суть работы модели выглядит так: человек вводит промт и получает ответ. Главное ее отличие от работы обычных программ в том, что *GPT способна создавать ответы в таком виде, как если бы их давал человек*.

Например, зададим ChatGPT задачу для пятого класса:

☞ **Реши задачу:** из автобуса на остановке вышло 6 пассажиров, а вошло 11. На следующей остановке вышло 8, вошло 9. Сколько пассажиров стало в автобусе, если вначале в автобусе было 24 пассажира?

Ответ модели, по сути, идентичен ходу мыслей обычного школьника (скопировано дословно).

Для решения этой задачи нужно проследить изменение количества пассажиров в автобусе на каждой остановке.

Изначально в автобусе было 24 пассажира. После первой остановки вышло 6 пассажиров, а вошло 11. Таким образом, количество пассажиров увеличилось на  $11 - 6 = 5$ .

$24 + 5 = 29$  пассажиров после первой остановки.

На следующей остановке вышло 8 пассажиров, а вошло 9. Снова посчитаем изменение количества пассажиров:

$29 - 8 + 9 = 30$  пассажиров после второй остановки.

Итак, после обеих остановок в автобусе стало 30 пассажиров.

Еще год-полтора назад GPT так не умела, а сегодня — запросто. Это происходит благодаря глубокому машинному обучению — непрерывному процессу, в который входит и обучение моделей новым

закономерностям, и расширение нейронных связей за счет увеличения программных мощностей, и колоссальная обратная связь с пользователями.

Чем больше мы общаемся с ChatGPT и чем больше реагируем, давая модели обратную связь, тем больше информации она получает и тем быстрее учится. Возникает тот самый эффект снежного кома.

«Выходит, нейросеть действительно умнеет и становится похожей на человека?» — спросите вы.

Честно — не знаем. Мы изучили массу докладов и лекций как ведущих ученых, IT-специалистов и разработчиков ИИ, так и нейропсихологов, нейрофизиологов и даже социологов. Среди экспертов нет единого мнения, поэтому вопрос остается открытым.

Мы приведем парочку фактов, а вы поразмышляйте и попробуйте сами сделать выводы.

*Факт первый.* Мы имеем дело уже не с программой, а с чем-то обладающим основами логики и анализа. И дальше навыки ИИ будут только расти. Все это видно по уже существующему прогрессу.

*Факт второй.* В техническом смысле нейросеть и не должна ничего «понимать» как человек. То есть ей для этого не нужны какие-то личностные или эмоциональные черты.

Почему? Потому что при глубоком машинном обучении сеть и не должна «очеловечиваться», она лишь ищет закономерности, закрепляет успешно подтвержденные и тем самым изучает, как работает человеческий язык. Для выполнения своих задач ей этого хватает.

Предполагается, что она лишь изучает вероятности и старается делать максимально «очеловеченный» вывод на основе наших данных. Например, что в такой-то связке слов и смыслов стоит употребить такие-то слова и смыслы и это даст лучшие результаты. Когда такое подтверждается много-много раз, сеть закрепляет это для себя как факт и делает частью стратегии.

Чтобы понять еще лучше, возьмем простой пример.

Есть некий Сергей Сергеевич, преподаватель истории в вузе. Раньше, рассказывая о временах Ивана Грозного (о которых он знает очень много), Сергей Сергеевич начинал издали: давал предпосылки, углублялся во второстепенное, долго подводил к сути и т. д.

Со временем он заметил, что такая стратегия не работает. Пока он «раскачивал тему», большинство студентов теряли интерес. Да, кое-кто слушал, но 95 % — зевали.

Тогда Сергей Сергеевич начал корректировать свои лекции. Он отмечал, какой материал студенты понимают, а какой — не очень, к чему они быстро проявляют интерес, а где начинают смотреть на часы. И вот так, шажками, он создал конструкцию лекции, которая и по объему, и по уровню знаний была интересна большинству.

И теперь, когда приходит время лекций об Иване Грозном, Сергей Сергеевич читает их в рамках этой конструкции. Она обкатана. Она работает и дает лучший результат.

Примерно то же самое происходит и с GPT, только таких Иванов Грозных у моделей миллиарды, и закономерности постоянно обновляются, оптимизируются, шлифуются.

*А теперь вопрос:* в разрезе повышения качества лекций так ли нам важно, какой Сергей Сергеевич человек? Дарит ли жене цветы, ворчит ли на молодежь и бьет ли котиков по лапкам? Пожалуй, не очень. В контексте преподавания он интересен нам только как некий носитель знаний и объект, который изучает закономерности конкретно для лекций. Всё.

Точно так же не обязательно, что GPT должна «очеловечиваться» или расти в каком-то личностном или эмоциональном плане. Главное, что модель умеет учиться и становится технически сильнее.

А вот наступит ли такой этап, когда, дойдя до определенного уровня развития, ИИ вдруг осознает себя как личность, или нет — это доподлинно не известно. Кто-то считает, что такое вероятно, кто-то только смеется. Так что не будем углубляться в эти дебри и просто пойдем дальше.

Итак, ChatGPT не только решает математические задачи. Модель открывает огромный пласт возможностей. С помощью нейросети можно делать что угодно:

- создавать опросы и анализировать текст;
- отвечать на вопросы и помогать с выбором на основе данных;
- придумывать диалоги и писать целые книги;
- программировать и разрабатывать новые алгоритмы;
- учиться языкам и сочинять поздравления;
- создавать контент для соцсетей и сайтов;
- играть в квестовые игры и искать логические ошибки;
- разрабатывать полезные меню и списки для саморазвития;
- находить лучшие аргументы для споров и писать сценарии;
- и многое, многое, многое другое...

Это очень сжатый перечень. А впереди нас ждут все те нюансы промтинга, которым посвящена книга. Мы вместе погрузимся в удивительный мир GPT и, что самое главное, научимся чувствовать себя там уверенно и комфортно.

## ДЛЯ КАКОЙ ВЕРСИИ GPT ПОДХОДИТ ЭТА КНИГА

Повторимся: изначально мы создавали книгу, которая никак не привязана ни к одной модели. И даже если завтра сменятся тренды и на место GPT придет какой-то новый игрок, все равно ничего не изменится. Книга все так же будет актуальна.

Почему? По двум причинам.

*Промтинг универсален.* Наша книга не о возможностях какой-то конкретной версии GPT или другой модели. И в ней нет набора из тысячи простых промтов, которые, честно говоря, любой человек может штамповать сотнями в час, если имеет минимальные навыки промтинга.

Эта книга о создании запросов вообще, о промт-инжиниринге (промтинге) как явлении и системе. Мы пока вообще не касались промтинга, но именно ему и посвящена вся книга.

Да, в книге будут примеры промтов из разных ниш. Но не в качестве основного наполнителя и самоцели, а лишь как инструменты для развития креативности и понимания путей. В основном же мы будем давать не рыбу, а удочки, чтобы вы сами ловили свою рыбу. А рыба и сегодня, и завтра будет ловиться примерно по тем же законам.

Почему? Потому что...

*ИИ развивается, но человек — статичен.* Возможно, когда-то будет достаточно коснуться приложения, и GPT 70-й версии просканирует ваш мозг, поймет все желания, вашу целевую аудиторию, нужный стиль текста, а потом сразу выдаст идеальный контент, от которого все упадут в экстазе.

Но пока оптимальная модель взаимодействия конкретна:

- человек спрашивает — ИИ отвечает;
- человек задает рамки — ИИ подгоняет ответ под них;
- человек уточняет — ИИ отвечает еще лучше (в идеале).

Всё. Иных путей взаимодействия человека и машины просто не существует. Да, ИИ может быть невероятно крутым и развиваться с бешеной скоростью, но все равно нейросеть должна ориентироваться на отстающего. На среднего человека. И никак иначе.

Всегда есть среднестатистический пользователь, который обладает определенным уровнем интеллекта. Если мы говорим о коммерческом

продукте, то ИИ все равно придется подстраиваться под того самого обычного человека, который просто хочет спросить и получить классный ответ.

Значит, все процессы взаимодействия с ИИ так и останутся примерно на одном уровне. Да, еще обязательно усилится точность и качество ответов в новых версиях, но это опять же говорит только о том, что будут расти качественные показатели машины, а не человека. Сам процесс общения не может ни сильно усложниться, ни упроститься.

*Почему не усложнится?* Потому что тогда потеряется массовая привлекательность продукта. Что-то слишком сложное не работает на широкую аудиторию. Да и прогресс — всегда про упрощение задач. Чем быстрее и легче, тем лучше, но никак не наоборот.

*Почему не упростится?* А уже некуда упрощать, промты сегодня и так максимально простые. Да, можно упростить какие-то конкретные детали, например вариант ввода голосом, возможность выбора стиля или библиотеки усилителей (как это сделано у нас в сервисе NeuroPanda (о нем расскажем чуть позже)).

Можно сделать надстройку к LLM с помощью обычного промтинга (например, А-блоки в NeuroPanda) или программирования (IT-архитектура под нужды конкретной компании), благодаря которым модели сразу будут знать, что вам нужен, скажем, текст лендинга в убедительном стиле.

Но сам по себе формат промтинга уже и так максимально приближен к реальности. Даже если вы будете просить не ИИ, а вполне себе живого специалиста написать вам какой-то конкретный текст, вы как минимум хотя бы раз тоже должны дать ему некие требования и вводные. Все то же самое, что и с ИИ.

Повторимся: эта книга не о версиях, а о мастерстве промт-инжиниринга вообще — как навыка, профессии и в чем-то даже искусства. В книге много практики, примеров, фишек и только самая полезная информация.

## КТО И ПОЧЕМУ НАПИСАЛ ЭТУ КНИГУ

Мы — сотрудники проекта «Панда-копирайтинг». Более десяти лет «ПК» создает тексты для бизнеса, рекламы, вирусного маркетинга, SMM, сайтов и т. д. Конкретно по контенту написали пять книг (это шестая). Создали стиль «убедительно-позитивный копирайтинг», в котором работают сегодня компании и профессиональные авторы.

Когда стало очевидно, что ChatGPT не очередной «халиф на час», а мощный инструмент, мы решили изучить ИИ пристальней. Начали с малого: стали пробовать первые стандартные промты, даже пытались создавать контент для наших второстепенных проектов.

К счастью, мы быстро увидели несостоятельность такого подхода. Стандартные подсказки работали кое-как, да и без четкой системы такая работа похожа скорее на лотерею. Когда-то повезет, а когда-то получается полная ерунда.

В итоге мы решили погружаться в тему полностью. Зимой 2023 года начали разработку сервиса NeuroPanda на базе GPT. Осенью того же года сервис начал работать и сейчас доступен по ссылке <https://neuropanda.app/>. Весь доступ к нейросети можно получить без VPN и прочих сложностей.

Кроме того, у него есть дополнительные сервисы наподобие библиотек графических и контентных подсказок, отдельная группа в соцсети для проектов «ПК» (<https://pandeon.pro/>), курсы (тоже бесплатные) и прочие интересные вещи.

«И при чем здесь книга?» — спросите вы.

Давайте зайдем чуть издалека, чтобы было понятнее.

Мы сразу решили, что точно не будем еще одним сервисом, просто транслирующим ChatGPT для конечного пользователя. Для подобной «легкотни» есть и оригинальный сайт OpenAI, и всевозможные боты в мессенджерах, например.

Конечно, и такой подход приносил бы деньги, однако на этом бы наше участие и заканчивалось. По сути, мы просто перепродавали бы доступ к оригиналу. Никакого творчества и романтики.

Мы сделали NeuroPanda другим. В основе его работы сейчас лежат два формата: А-блок и чаты.

*А-блоки* — разделы с блоками, преднастроенные определенными промтами для конкретных задач. Например, А-блоком является крупный раздел с блоками для написания большой экспертной статьи.

Мы разбили задачу на шаги, в ходе которых вы можете собрать все элементы (заголовки, темы, идеи вообще и разделов, списки, кластеры и т. п.) так, как вам удобно, а затем на основе всего этого сделать добротную статью.

Если делать так самому просто в режиме чат-подсказки (пусть и многоуровневой, мощной), то все куда более сложно и непредсказуемо: слишком много вводных, слишком многое нужно учитывать, слишком велика вероятность запутать ИИ и нарваться на галлюцинации (это явление рассмотрим позже).

*Чаты* представляют собой как «чистый» ChatGPT без преднастроек, так и специализированные чаты для разных задач. Один, например, «заточен» на обучение языкам, другой поможет лучше находить логические ошибки или готовиться по конкретной теме.

А вот теперь о том, при чем здесь книга.

Процессы такого уровня — колоссальный труд. Нам пришлось не просто изучить подсказки изнутри и потратить много-много миллионов токенов («валюта» и мера объема GPT) в процессе работы. Нам волей-неволей пришлось:

- открыть массу фишек и закономерностей;
- сделать тысячи ошибок и понять их причину;
- найти интересные стратегии промтинга;
- обкатать сотни вариантов структуры и запросов;

- разочароваться в поверхностных промтах;
- прокачать креативность и увидеть новые пути;
- и т. д. и т. п.

В итоге мы стали сильными практиками промт-инжиниринга и накопили такой объем полезного *практического* багажа, который так и просился для отдельной книги. Именно поэтому она родилась.

P. S. Кстати, сейчас, поняв и освоив стандартный промтинг, мы уже вплотную подошли к более серьезному этапу, а именно — к созданию собственных LLM для точечных задач. Например, к написанию ИИ текста с определенной стилистикой.

Конечно, речь идет о надстройках на предобученные модели, а не о подготовке LLM с нуля, но в нашем случае такого и не требуется. Тонкое обучение — тоже не фунт изюма, но, надеемся, справимся и здесь.