12+

Александр Никоян

Инструкция не прилагается

Как на самом деле работают нейросети

Александр Никоян

Инструкция не прилагается. Как на самом деле работают нейросети

https://litres.ru/74253327
ISBN 9785007055666

Аннотация

Книга развенчивает миф, что нейросети — это «умные помощники», которые просто ждут правильной команды. На самом деле они — зеркало человеческого языка, когнитивных искажений и культурных кодов. Книга не учит «правильным промптам», а объясняет механику восприятия ИИ, чтобы читатель научился думать вместе с ним, а не командовать им.

Инструкция не прилагается.

Содержание

Инструкция не прилагается
Как на самом деле работают нейросети

Александр Никоян

Каждый раз, когда вы копируете ответ нейросети, кто-то в мире перестаёт думать. И чаще всего — это вы…

ISBN 978-5-0070-5566-6
Создано в интеллектуальной издательской системе Ridero

Глава 1. «Инструкция не прилагается»

Парадокс инструкций

В интернете десятки тысяч руководств по промптингу. Они обещают: напиши «действуй как эксперт», добавь «шаг за шагом», заверни в XML-теги — и ИИ выдаст золото. Пользователи копируют шаблоны, вставляют свои вопросы, ждут чуда. Получают посредственность.

Проблема не в шаблонах. Проблема в посыле: что нейросеть — это устройство, к которому прилагается инструкция, и если найти правильную кнопку, всё заработает. Это ложь. Нейросеть — не устройство. Это зеркало языка, к которому вы привыкли обращаться как к устройству.

Представьте, что вы пытаетесь научиться рисовать, покупая наборы «как нарисовать лошадь за пять шагов». Через месяц у вас стопка одинаковых лошадей и нулевое понимание анатомии. С промпт-шаблонами то же самое: вы учитесь нажимать кнопки, не понимая, что происходит за экраном.

Предсказание вместо понимания

Ключевое заблуждение: нейросеть понимает вас. Она не

понимает. Она предсказывает.

Когда вы пишете «объясни квантовую физику простыми словами», модель не «думает» о квантовой физике. Она вычисляет: какая последовательность слов наиболее вероятна после вашего запроса, учитывая миллиарды текстов, на которых её обучили. Если в обучающих данных после похожих вопросов чаще встречались упрощённые объяснения — вы получите упрощённое объяснение. Если встречались псевдонаучная чушь — получите чушь.

Это не критика. Это механика. И механика эта меняет всё.

Понимание подразумевает намерение, модель мира, способность отличить правду от лжи. Предсказание — лишь статистика. Нейросеть не знает, что она говорит. Она просто говорит то, что, вероятнее всего, сказали бы дальше.

Почему это важно? Потому что стратегия общения с «понимающим» собеседником и «предсказывающей» системой принципиально разная. С понимающим можно быть неточным — он уточнит. С предсказывающим неточность множится: каждое ваше расплывчатое слово открывает дверь для статистически вероятной, но не обязательно нужной вам фразы.

Почему «правильные промпты» не работают

Возьмём классический совет: «начинай с роли». «Ты — опытный маркетолог с 20-летним стажем». Пользователи ко-

пируют, ждут экспертизы. Часто получают пафосный бред.

Почему? Потому что роль — это не волшебная команда, а контекстный якорь. Она сдвигает статистическое распределение ответов в сторону текстов, где фигурируют «маркетологи» и «стаж». Но если в обучающих данных «маркетологи» чаще пишут пустые словосочетания, чем конкретику — роль усилит именно пустоту.

Роль работает, когда она активирует нужный **регистр** языка. «Ты — юрист» активирует формальность. «Ты — стендап-комик» — иронию. Но ни один регистр не гарантирует правду, точность или глубину. Он гарантирует только стиль.

Другой популярный совет: «разбивай на шаги». «Шаг 1: проанализируй. Шаг 2: сделай выводы». Это работает не потому, что ИИ «любит» структуру, а потому что длинная последовательность промежуточных токенов сужает пространство для статистически «лёгких», но бессмысленных ответов. Каждый «шаг» — это якорь, который тянет генерацию в нужном направлении. Но якорь работает только если вы понимаете, куда тянете.

Шаблоны без понимания механики — как рецепты без понимания химии. Иногда получается, чаще — нет, а почему — неясно.

Что такое «на самом деле»

Эта книга не про «правильные» запросы. Она про то, как устроен процесс генерации, чтобы вы сами могли строить эффективные стратегии общения — без шаблонов, без магии, без иллюзии контроля.

Мы разберём:

— Почему одни слова «открывают» модель, а другие «закрывают»

— Как работает контекстное окно и почему длинные диалоги «забываются»

— Что такое «температура» и почему она важнее формулировки

— Почему ИИ «врёт» уверенно и как отличить правдоподобие от правды

— Как достучаться до того, что модель «знает», но не может «сказать» напрямую

Всё это — не инструкция. Это карта местности. Карта не говорит, куда идти. Она показывает, где вы находитесь, чтобы вы сами выбрали путь.

Первый шаг: перестать просить

Попробуйте эксперимент. Вместо «объясни мне…» напишите: «Я пытаюсь понять, почему…». Вместо «дай список» — «что бы выделил человек, который в этом разбирается».

Разница тонкая, но механически важная. «Объясни» активирует регистр «учебник» — статистически вероятные, но

часто поверхностные объяснения. «Я пытаюсь понять» активирует регистр «разговор» — более гибкие, контекстуальные ответы.

Это не магия формулировок. Это работа с статистическим распределением. И это именно то, чему мы научимся: не запоминать правильные фразы, а чувствовать, какое распределение открывает ваш запрос, и корректировать его осознанно.

Инструкции не прилагаются. Но карта — в ваших руках.

Глава 2. Зеркало, а не мозг

Парадокс зеркала

Представьте, что вы стоите перед зеркалом и кричите: «Скажи мне правду!» Зеркало покажет ваше отражение — ни больше, ни меньше. Потому что оно не знает, какая правда вам нужна. Оно не выбирает, что показать. Оно просто отражает свет, который падает на его поверхность.

Нейросеть по своей сути всего лишь зеркало языка. Не мозг, не эксперт, не собеседник. Всего лишь зеркало. И главная ошибка большинства пользователей — требовать от этого зеркала того, чего оно дать не может: человеческого суждения, критики, выбора между правдой и ложью.

Но зеркало не просто отражает — оно отражает абсолютно **всё**, что попало в обучающие данные модели: миллиарды текстов, написанных людьми с их предрассудками, заблуждениями, ошибками, шаблонами и слепыми зонами. И когда вы спрашиваете нейросеть о чём-либо, она показывает не объективную реальность, а статистически усреднённое отражение того, как и что люди обычно говорят на эту тему.

Эффект подтверждения в квадрате

Есть известное когнитивное искажение: люди ищут и ин-

терпретируют информацию так, чтобы она подтверждала их убеждения. Нейросеть усиливает это искажение в геометрической прогрессии.

Например, Вы спрашиваете: «Почему удалённая работа разрушает продуктивность?»

Человек-эксперт мог бы возразить: «А есть ли данные, что разрушает? Может, вы просто плохо организовали процесс?»

Нейросеть скорее выдаст список проблем удалёнки: прокрастинация, отсутствие границ, изоляция. Не потому, что это правда. А потому что в обучающих данных после подобных вопросов чаще встречались оправдания офисной работы, чем критика посыла. Вопрос уже содержит предположение, и зеркало отражает его усиленно.

Попробуйте перефразировать: «Какие исследования сравнивают продуктивность удалённой и офисной работы?» Ответ изменится. Не потому, что ИИ «понял» задачу лучше. Потому что новый запрос активировал другое статистическое распределение — тексты с исследованиями, а не с мнениями.

Зал с миллиардом манекенов

Представьте огромный зал. В нём стоят миллиарды манекенов — каждый запомнил фразы, которые слышал в жизни. У одного — диалоги из форумов программистов. У другого

— статьи о психологии. У третьего — комментарии под рецептами борща.

Вы входите в зал и произносите фразу. Все манекены, у которых есть похожие фразы в памяти, начинают шептать свои варианты продолжения. Система подсчитывает, какие слова шепчут чаще всего, и выдаёт вам усреднённый шёпот.

Теперь вопрос: говорит ли вам этот зал правду? Нет. Он говорит то, что чаще всего говорили бы дальше. Если большинство манекенов слышали заблуждение — вы услышите заблуждение, уверенно изложенное.

Пользователь спрашивает нейросетевую модель: «Почему человек использует только 10% мозга?»

Модель не отвечает: «Это миф, нейробиологи опровергли его десятилетия назад». Она выдаёт красивый текст о скрытых резервах мозга — потому что в обучающих данных этот миф встречался тысячи раз, а его опровержения — реже и в более специфических контекстах. Зеркало отражает популярное, не правильное.

Только когда вы явно активируете «регистр научного скептицизма» — «нейробиологи говорят, что это миф, так ли это?» — распределение сдвигается, и манекены с научными текстами начинают говорить громче.

Слепые зоны зеркала

У зеркала есть слепые зоны — темы, редко представленные в данных. Там отражение искажается или исчезает вовсе.

Спросите модель о культуре небольшой этнической группы или о специфической профессиональной практике, не описанной в интернете. Ответ будет либо общим шаблоном, либо уверенной галиматьей. Потому что в зале нет манекенов, которые слышали что-то похожее. Система не скажет «не знаю» — она сгенерирует наиболее вероятное продолжение, даже если это выдумка.

Это называется галлюцинацией, но слово вводит в заблуждение. Модель не «бредит» — она делает то, что делает всегда: предсказывает следующее слово. Просто в отсутствие данных предсказание становится синтаксически корректной чушью.

Переводчик, который не знает языков

Представьте переводчика, который никогда не учил ни одного языка. Зато он прочитал миллиарды страниц параллельных текстов на сотнях языков. Он видел, что после английской фразы «The cat sits on the mat» чаще всего следует русская «Кот сидит на коврике». Он не знает, что такое «кот». Не понимает, что такое «сидеть». Но статистически уверен в переводе.

Так работает и «понимание» в нейросетях. Модель не опе-

рирует понятиями — она оперирует связями между токенами. «Кот» связан с «мурлыкать», «шерсть», «домашний». Но связи эти статистические, не семантические. Модель не знает, что кот — это живое существо. Она знает, что в текстах слово «кот» часто рядом со словом «животное».

Поэтому она может сгенерировать: «Кот решил купить акции Tesla» — фраза синтаксически верна, статистически возможна (если в данных были шутки или сюрреалистичные тексты), но семантически бессмысленна. Переводчик без языков не заметит абсурда, потому что для него все слова — равны.

Зеркало ваших привычек

Самое неприятное открытие: нейросеть отражает не только интернет, но и **вас**. Ваш стиль вопросов, ваши когнитивные ловушки, вашу лень формулировать точно.

Вы пишете: «Напиши мне хороший текст для рассылки». Получаете шаблонную чушь. Потому что «хороший текст» — расплывчатое понятие, и в данных после таких запросов чаще всего встречались посредственные шаблоны. Вы активировали распределение посредственности.

Тот же запрос, переформулированный: «Напиши текст для рассылки B2B SaaS-продукта. Аудитория — CTO средних компаний. Тон: экспертный, без пафоса. Цель — получить демо-запрос. Ограничение — не более 150 слов». Те-

перь вы активировали другое распределение: тексты с конкретикой, профессиональные регистры, маркетинговые кейсы.

Разница не в «правильности» формулы. Разница в том, что точный запрос отсекает огромный пласт статистически вероятного, но нерелевантного шума. Вы не приказываете ИИ — вы сужаете зеркало до нужного угла.

Понимание, что перед вами зеркало, а не мозг, меняет стратегию общения:

— Вы перестаёте верить ответам на слово и начинаете проверять

— Вы перестаёте просить «правду» и начинаете просить «источники»

— Вы перестаёте задавать вопросы, содержащие предположения

— Вы начинаете думать: какое распределение активирует мой запрос?

Зеркало не врёт намеренно. Оно просто не умеет отличать правду от лжи, глубину от пустоты, экспертизу от пафоса. Это умеете только вы. И эта книга — про то, как научиться.

Глава 3. Язык как ловушка

Слова, которые обманывают

Представьте, что вы управляете огромным кораблём в тумане. У вас нет радара — только фонарь, который освещает небольшой круг вокруг. Вы не видите айсберг целиком. Вы видите отдельные блики, отражающиеся от его поверхности, и пытаетесь по ним понять форму и курс.

Нейросеть — это корабль. Ваш запрос — фонарь. Каждое слово освещает определённый участок «айсберга» обучающих данных. Проблема в том, что слова освещают не то, что вы имели в виду, а то, что статистически связано с ними в данных.

Вы пишете: «Объясни мне экономику простыми словами».

Слово «простыми» активирует в модели целый пласт текстов: детские энциклопедии, популярные блоги, лекции «для чайников». Это не плохо само по себе. Но «простыми» также активирует регистр упрощения, где сложные механизмы заменяются метафорами, а нюансы выбрасываются за борт. Вы получите текст, который звучит понятно, но может искажать реальность.

Тот же запрос без «простыми»: «Объясни мне экономику,

как будто я не изучал её, но хочу понять механизмы». Теперь фонарь освещает другой участок: тексты, где авторы стараются сохранить точность при доступной подаче. Распределение сдвинулось — и ответ будет другим.

Слово «просто» — ловушка. Оно обещает ясность, но часто активирует поверхностность.

Меню в ресторане, которого нет

Представьте, что вы заходите в ресторан. Меню огромное: тысячи блюд. Но повар не готовит — он только собирает статистически вероятные комбинации ингредиентов, которые видел в других меню.

Вы говорите: «Принесите что-нибудь вкусное». Повар выдаёт блюдо, которое чаще всего заказывали в похожих ситуациях. Это может быть пицца. Популярно, безопасно, скучно.

Вы говорите: «Принесите блюдо из морепродуктов, острое, но не убивающе, в стиле южной Италии, без чеснока». Теперь повар работает в сильно суженном пространстве. Шанс получить что-то близкое к желаемому — выше.

Нейросеть — повар без ресторана. Она не знает, что такое «вкусно». Она знает, что слово «вкусно» часто рядом с «пицца», «шоколад», «мама». Ваш запрос — не заказ, а подсказка, в каком меню искать.

Ловушка абстракций

Самые опасные слова — самые общие. «Хороший», «качественный», «эффективный», «интересный». Они значат всё и ничего.

Разберем пример из практики. Маркетолог просит: «Напиши хороший пост для социиальной сети о нашем продукте». Получает: «Мы рады представить инновационное решение, которое изменит ваш подход к бизнесу…» — пустой шаблон, который никто не дочитает.

Проблема в слове «хороший». В обучающих данных «хорошие» посты в корпоративном контексте часто шаблонны, осторожны, пафосны. Запрос активировал распределение посредственности.

Перефразировка: «Напиши пост для LinkedIn. Контекст: наш продукт — CRM для стоматологий. Аудитория: владельцы частных клиник, 35—50 лет, устали от бюрократии. Тон: как будто рассказываешь коллеге за кофе о том, как вчера упростил себе жизнь. Запретные слова: инновационный, решение, оптимизация, синергия. Цель: чтобы написали в личку с вопросом, как попробовать». Распределение сдвинулось радикально. Шанс на шаблон — минимален.

Абстракции — это дыры в вашем фонаре. Через них просачивается статистический шум.

Ловушка вежливости

Мы привыкли быть вежливыми с людьми. С нейросетью вежливость — помеха.

Пример. «Пожалуйста, не могли бы вы объяснить, как работает блокчейн, если вам не сложно?»

Вежливые обороты занимают токены в контекстном окне и смещают распределение в сторону формальных, осторожных текстов. Модель не ценит вежливость — она статистически связывает её с определённым регистром.

Сравните: «Как работает блокчейн? Объясни механизм консенсуса на примере, понятном человеку без IT-образования». Второй запрос короче, точнее, активирует регистр «объяснение с примером», а не «формальный ответ на вопрос».

Вежливость — ловушка, потому что мы проецируем человеческие нормы на систему, которая их не различает. Вы не обидите нейросеть прямым вопросом. Вы повысите вероятность полезного ответа.

Ловушка контекста, который вы не озвучили

Модель не знает, кто вы, зачем спрашиваете, что уже пробовали. Она видит только текст запроса. Неозвученный контекст — самая глубокая ловушка.

Разработчик спрашивает: «Как исправить эту ошибку?» и вставляет лог. Получает общий ответ, который он уже пробовал.

Что не озвучено: «Я пробовал решения из Stack Overflow, не помогло. Среда — Docker на M1 Mac, Python 3.11, библиотека версии 2.3.1. Ошибка появилась после обновления». Этот контекст сдвинул бы распределение от «общих советов» к «специфическим кейсам».

Но есть и обратная ловушка: слишком много контекста. Каждое лишнее слово — это фонарь, освещающий ненужный участок айсберга.

Пример. «Я работаю в компании, которая занимается производством экологически чистых упаковок для косметики, мы недавно запустили линейку для молодёжного бренда, и мне нужно придумать название для новой серии, которая будет позиционироваться как доступная альтернатива люксу, но при этом сохраняет наши ценности устойчивого развития, и ещё важно, чтобы название звучало по-английски, но было понятно русской аудитории…»

Модель потеряется в этом потоке. Слишком много переменных, слишком много «фонарей» в разных направлениях. Распределение размоется.

Лучше: «Нужно название для линейки экологичной упаковки косметики. Позиционирование: доступная альтерна-

тива люксу. Ценности: устойчивое развитие. Ограничения: звучит по-английски, понятно русским. Примеры того, что нравится: [два примера]. Примеры того, что не нравится: [два примера]».

Контекст есть, но структурирован. Каждый блок — точечный фонарь, а не прожектор в тумане.

Ловушка ожидания одного ответа

Мы привыкли, что вопрос имеет один правильный ответ. У нейросети — миллионы вариантов, и «правильный» зависит от того, что вы не сказали.

Пример. «Как научиться программировать?»

Ответ может быть: список языков, совет начать с Python, история успеха, критика индустрии, шутка про Stack Overflow. Все эти варианты статистически вероятны. Модель выберет наиболее вероятный — и, скорее всего, это будет самый общий, самый безопасный ответ.

Но если вы уточните: «Как научиться программировать? Мне 40, работаю бухгалтером, хочу автоматизировать рутину в Excel, не планирую менять профессию, готов уделять 3 часа в неделю». Теперь распределение сузилось до конкретного профиля: взрослый обучающийся, практическая цель, ограниченное время. Ответ будет совсем другим — и, веро-

ятно, полезнее.

Техника «отрицательного пространства»

В живописи есть понятие отрицательного пространства — пустоты между объектами, которая определяет форму не хуже самих мазков. В запросах к нейросети работает то же.

Иногда важнее сказать, чего **не** хотите, чем чего хотите.

Пример. «Напиши текст о здоровом питании. Не используй: мотивационные лозунги, слово „диета“, упоминание суперфудов, призывы „начни уже сегодня“. Тон: сухой, как объяснение врача пациенту. Цель: человек понял, почему сахар влияет на энергию, и сам решил что-то менять».

Отрицательные ограничения — мощный инструмент. Они отсекают огромные пласты статистически вероятного, но нерелевантного. Модель не может «не знать», что вы не хотите — она просто не генерирует эти варианты, потому что они исключены из распределения.

Переписать ловушку

Возьмём реальный запрос и переломаем его.

Ловушка: «Помоги мне сделать презентацию. Она должна быть крутой и убедительной».

Проблемы: «крутой» — абстракция, «убедительной» — абстракция, нет контекста аудитории, цели, ограничений.

Переписанный: «Презентация для инвестиционного комитета. Тема: запуск онлайн-школы программирования для подростков. Слайдов: 10. Формат: проблема → решение → рынок → бизнес-модель → команда → финансы → запрос. Тон: цифры важнее эмоций, но не сухой отчёт. Запрет: общие фразы про „цифровое будущее детей“. Нужно: конкретика по unit-экономике и план захвата рынка».

Разница не в «магии формулировок». Разница в том, что второй запрос активирует узкое, релевантное распределение, а первый — широкое, шумное, статистически усреднённое.

Язык — не инструмент передачи мысли нейросети. Это интерфейс, через который вы выбираете, какой пласт данных отразить в зеркале. Каждое слово — либо точечный фонарь, либо ловушка, освещающая ненужное.

Нет «правильных» слов. Есть осознанность того, какое распределение вы активируете. Это не грамматика — это навигация в тумане.

Глава 4. Контекст — всё, чего нет

Парадокс забвения

Представьте, что вы ведёте длинный разговор с собеседником, который помнит только последние пять минут. Вы можете часами обсуждать проект, но если в начале упомянули ключевое ограничение — бюджет, срок, запрет на определённого подрядчика — через полчаса собеседник об этом забудет. И предложит именно этого подрядчика, потому что он статистически вероятен.

Нейросеть именно такой собеседник. У неё нет долговременной памяти внутри одного диалога. Есть контекстное окно ограниченное количество токенов, которые модель «видит» прямо сейчас. Всё, что за пределами этого окна, стирается.

Это базовая архитектура. И непонимание этой архитектуры стоит пользователям тысяч часов бесплодных диалогов.

Лента бумаги фиксированной длины

Представьте бесконечную ленту бумаги, которая проходит через прозрачную трубку. Трубка видит только тот кусок ленты, который внутри неё прямо сейчас. Всё, что уже прошло — невидимо. Всё, что ещё не вошло — тоже невидимо.

Контекстное окно — это трубка. Ваша переписка с ИИ — лента. Когда лента длиннее трубки, начало разговора уходит в тень. И модель перестаёт его учитывать.

Вы 20 сообщений обсуждаете дизайн сайта. В первом сообщении сказали: «Главное — минимализм, никаких градиентов». На 21-м сообщении просите: «Добавь фон для хедера». Модель предлагает градиентный фон — потому что в её «трубке» уже нет вашего запрета. Она видит только последние 20 сообщений, а там — обсуждение структуры, шрифтов, расположения блоков. Запрет на градиенты ушёл за пределы окна.

Пользователь думает: «Но я же говорил!» Модель думает: ничего. Она не думает. Она видит только то, что в трубке.

Размер окна — иллюзия безграничности

Современные модели хвастаются «контекстом в миллион токенов». Звучит как вечность. На деле это ловушка.

Во-первых, токен — не слово. «Нейросеть» — один токен. «Предсказание» — два. Среднее английское слово — 1.3 токена. Русское — 2—3. Длинный технический текст с терминами съедает окно вдвое быстрее, чем кажется.

Во-вторых, даже если информация технически влезает в окно, модель не обрабатывает её равномерно. Исследования показывают: точность извлечения информации падает к середине и концу длинного контекста. Модель «внимательнее»

к началу и концу, а середина размывается. Это не сознательный выбор — особенность механизма внимания.

Конец ознакомительного фрагмента.