

Не верь нейросети на слово

Как проверять ответы ИИ
и не попадаться на ошибки



Юрий Тер

Юрий Тер

Не верь нейросети на слово: Как проверять ответы ИИ и не попадаться на ошибки

<https://litres.ru/74385848>

SelfPub; 2026

Аннотация

«Не верь нейросети на слово» — практическая книга о том, как пользоваться искусственным интеллектом без слепого доверия и без страха перед новой технологией. Почему нейросеть может уверенно сообщить несуществующий факт, сослаться на вымышленный источник, неверно пересказать настоящий документ или выдать устаревшие данные за актуальные? Чем генеративная модель отличается от энциклопедии и обычного поисковика — и в каких ситуациях её ответ обязательно нужно проверять?

На реальных примерах разбираются ошибки ИИ в поиске информации, медицине, документах, отзывах, изображениях, голосе и видео. Читатель узнает, как быстро проверять факты и цитаты, отличать первоисточник от пересказа, оценивать свежесть информации, распознавать цифровой обман и не поддаваться на срочность мошеннических сообщений. Отдельно

показано, где нейросеть действительно полезна и как превратить её из «последней инстанции» в эффективного помощника.

Содержание

Вступление	5
Помощник, которому хочется верить	5
ЧАСТЬ I. ЧТО ТАКОЕ НЕЙРОСЕТЬ И	9
ПОЧЕМУ ОНА МОЖЕТ ОШИБАТЬСЯ	
Глава 1. Нейросеть, энциклопедия и	9
поисковик — три разных инструмента	
Энциклопедия: готовая статья	10
Поисковик: список путей к информации	11
Нейросеть: новый текст	12
А если нейросеть подключена к интернету?	13
Короткое сравнение	14
Конец ознакомительного фрагмента.	16

Юрий Тер

Не верь нейросети на слово: Как проверять ответы ИИ и не попадаться на ошибки

Вступление

Помощник, которому хочется верить

Представьте, что у вас появился помощник, который почти никогда не устает. Он может за несколько секунд объяснить незнакомый термин, составить план поездки, переписать письмо, придумать идею, разобрать документ и помочь начать работу в области, о которой вы вчера почти ничего не знали.

Он отвечает спокойно и грамотно. Не раздражается, если вы переспросили. Не просит назначить встречу через неделю. Может объяснить одну и ту же мысль пятью способами — пока вы наконец не поймёте.

Очень быстро возникает естественное ощущение: перед вами почти универсальный эксперт.

Именно здесь начинается главная ошибка новичка.

Нейросеть — не электронная энциклопедия, в которой лежит готовая карточка с правильным ответом. И не поисковик в привычном смысле, который прежде всего показывает страницы, где ответ можно проверить. Генеративная модель создаёт ответ. Иногда на основе собственных параметров, иногда с помощью подключённого поиска или загруженных документов, но итоговый текст всё равно формируется моделью.

Это различие кажется техническим, пока не приходится платить за ошибку.

В 2025 году на туристическом сайте Tasmania Tours появилась статья о горячих источниках в Тасмании. Среди настоящих мест были описаны Weldborough Hot Springs — привлекательная лесная локация с минеральной водой и красивыми изображениями. Люди стали туда ехать. Одна группа из 24 путешественников специально сделала крюк. На месте выяснилось, что горячих источников там нет. [6]

История почти смешная: лишняя дорога, потраченное время, испорченный кусок отпуска. Но тот же механизм выглядит иначе, если ответ касается лекарства, судебного документа или крупного перевода денег.

В 2024 году был опубликован клинический случай пациента, который использовал ChatGPT как часть самостоя-

тельной оценки тревожных симптомов и примерно на сутки отложил обращение за неотложной помощью. В больнице рабочим диагнозом стала транзиторная ишемическая атака. [7] В 2025 году врачи описали другой случай: мужчина около трёх месяцев заменял поваренную соль бромидом натрия после разговора с ChatGPT об альтернативах хлориду натрия и оказался в больнице с бромизмом. [8]

Во всех этих историях важен один переход: текст на экране превратился в основание для действия.

Поэтому эта книга не написана против нейросетей. Напротив: генеративный ИИ — один из самых полезных новых инструментов для работы с текстом, идеями и информацией. Но пользоваться им безопаснее и эффективнее, когда вы понимаете, чем он является — и чем не является.

Хорошо сформулированный ответ и правильный ответ — не одно и то же.

NIST называет уверенно сформулированное ошибочное или ложное содержание генеративных систем конфабуляцией; в повседневной речи чаще употребляют слово «галлюцинация». NIST подчёркивает, что такие ошибки естественно связаны с тем, как генеративные модели строят статистически правдоподобные продолжения. [1]

В 2025 году исследователи OpenAI предложили ещё одну важную деталь: галлюцинации поддерживаются не только самим предобучением, но и системой оценки, в которой модели часто выгоднее угадать, чем признать «я не знаю». [26]

Это не единственное объяснение; научные обзоры рассматривают причины по всему жизненному циклу модели — от данных и обучения до генерации ответа, контекста запроса и подключения внешних источников. [27]

Первая задача этой книги — убрать магию. После первых глав у вас должно сложиться простое представление: нейросеть не хранит внутри идеальную энциклопедию мира и не является просто улучшенным Google или Яндексом. Это другой тип инструмента.

Вторая задача — научиться использовать этот инструмент без двух крайностей. Не верить ему на слово. Но и не превращаться в человека, который боится каждого ответа ИИ.

Главный вопрос перед важным действием: «Откуда я знаю, что это правда?»

ЧАСТЬ I. ЧТО ТАКОЕ НЕЙРОСЕТЬ И ПОЧЕМУ ОНА МОЖЕТ ОШИБАТЬСЯ

Глава 1. Нейросеть, энциклопедия и поисковик — три разных инструмента

Когда человек впервые открывает чат с нейросетью, он обычно уже знает, как пользоваться энциклопедией и поиском. Поэтому мозг автоматически пытается поместить новый инструмент в одну из знакомых категорий.

«Наверное, это огромная энциклопедия, которая умеет разговаривать». Или: «Это поисковик, только он сразу пишет готовый ответ». Оба сравнения полезны как отправная точка — и оба становятся опасными, если принять их буквально.

Энциклопедия: готовая статья

Энциклопедия хранит подготовленные материалы. Кто-то написал статью, кто-то её редактировал, а читатель получает уже существующий текст. Ошибки возможны, сведения могут устареть, качество редакции бывает разным — но принцип понятен: материал существовал до вашего вопроса.

Поисковик: список путей к информации

Классический поисковик работает иначе. Вы вводите запрос, а система находит документы и страницы, которые считает релевантными. Главный продукт поиска — не новое объяснение, а путь к источникам. Пользователь может открыть несколько страниц, увидеть автора, дату, контекст и сравнить версии.

Современный поиск всё чаще добавляет генеративные сводки, а нейросети всё чаще подключают веб-поиск. Поэтому граница между продуктами стала менее заметной. Но два механизма всё равно полезно различать: поиск извлекает источники, генеративная модель формирует новый ответ.

Нейросеть: новый текст

Большая языковая модель обучается на огромных массивах текста и учится предсказывать продолжение последовательности. В упрощённом виде она много раз решает задачу: «какой следующий фрагмент текста наиболее подходит здесь?» Из миллиардов таких задач постепенно возникает способность писать связные ответы, объяснять, переводить, обобщать и следовать инструкциям. [26]

Поэтому полезная мысленная модель звучит так: энциклопедия хранит статью; поиск находит источники; нейросеть конструирует ответ.

Это не означает, что модель «ничего не знает». В её параметрах действительно закодировано огромное количество статистических закономерностей и сведений. Но это не картотека, где для каждого факта есть отдельная карточка с пометкой «проверено». Модель может воспроизводить известный факт, обобщать несколько знаний, сделать вывод — или построить правдоподобную комбинацию, которой в реальности не было.

А если нейросеть подключена к интернету?

Тогда перед нами гибрид. Сначала система может найти документы, затем языковая модель прочитает фрагменты и сформирует ответ. Это обычно повышает надёжность свежих фактов, но не отменяет новые точки ошибки: найден может быть плохой источник, документ можно понять неверно, важное исключение можно пропустить, а несколько сайтов могут копировать одну и ту же ошибку.

Поэтому наличие веб-поиска меняет вопрос «откуда модель взяла данные?», но не отменяет вопрос «подтверждает ли источник именно этот вывод?»

Короткое сравнение

Инструмент

Главная задача

Что получает пользователь

Главный риск

Энциклопедия

Дать подготовленную справочную статью

Существующий отредактированный материал

Ошибки редакции и устаревание

Поисковик

Найти релевантные страницы

Список источников и фрагментов

Плохой источник или неверная оценка результата

Нейросеть

Сформировать ответ по контексту

Новый связный текст

Правдоподобный, но неверный вывод

Запомнить достаточно одной фразы: если поисковик прежде всего помогает найти, то нейросеть прежде всего по-

могает сформулировать. Смешивать эти роли удобно — но только когда вы понимаете, какой этап выполняется сейчас.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.