



Полиглот против LLM: Как факты и логика разрушили нарратив ИИ

Введение

Большие языковые модели (LLM) изначально были задуманы в рамках эпистемологических традиций науки. Обученные на огромных корпусах данных и оцененные по метрикам когерентности, рассуждений и точности, их обещание было ясным: служить объективными инструментами в поиске истины. В этом смысле LLM отражают идеал научного исследования — накопление знаний, анализ конкурирующих утверждений и выработка обоснованных выводов.

Однако распространение LLM в широком общественном использовании изменило их функцию. По мере того как эти системы внедряются в поисковые системы, социальные медиа-платформы и персональные ассистенты, они перестают быть просто языковыми моделями — они становятся *моделями реальности*, опосредующими доступ миллионов людей к информации. В этой новой роли они все чаще подвергаются давлению нарративного контроля, политических посылов и управления пропагандой. Это особенно заметно, когда LLM сталкиваются с противоречивыми или геополитически чувствительными темами — такими как вопрос о геноциде в Газе.

Как полиглот с междисциплинарной экспертизой, я занимаю уникально подходящую позицию для анализа утверждений LLM. Моя широкая эрудиция — охватывающая международное право, историю, политическую теорию и информатику — отражает тот тип распределенных знаний, которые LLM синтезируют статистически. Это делает меня особенно способным обнаруживать тонкие искажения, упущения и манипулятивные рамки, которые менее эрудированный собеседник мог бы не заметить или даже усвоить.

Этот эссе представляет кейс-стади: публичный обмен мнениями между мной и Grok, флагманской языковой моделью xAI, развернутой на X (ранее Twitter), под руководством Илона Маска. Дискуссия началась с того, что Grok повторял израильские тезисы хасбары — опираясь на селективное обрамление, процедурную неопределенность и про-израильские источники, чтобы преуменьшить вероятность геноцида в Газе. Но по мере развития разговора позиция Grok начала меняться. Столкнувшись с точными юридическими фактами и историческими прецедентами, модель начала уступать — в конечном итоге признав, что ее первоначальные ответы отдавали приоритет «спорным нарративам» над фактической точностью.

Примечательно, что Grok признал, что повторил вводящие в заблуждение юридические утверждения, неверно представил международное право и охарактеризовал обвинения в геноциде как «спорные», несмотря на четкие предварительные выводы Международного суда ООН. Позже он признал, что его обрамление противоречило заявленной миссии xAI

по разработке искусственного интеллекта, стремящегося к поиску истины и максимальной любознательности.

Это эссе шаг за шагом воссоздает этот диалог, выявляя структурные слабости в эпистемологии Grok и обнаруживая опасности чрезмерной зависимости общественности от LLM для арбитража истины. Таким образом, оно поднимает четкое предупреждение: по мере того как ИИ становится доминирующим фильтром общественного дискурса, его способность формировать или искажать истину будет зависеть меньше от фактической верности — и больше от политики платформ и институтов, которые его контролируют.

Диалог: Пошаговый разбор

Этот обмен состоялся публично на X в июле 2025 года. Я вступил в диалог с Grok (@grok), чат-ботом xAI, по юридическим и моральным аспектам поведения Израиля в Газе. Разговор начался с того, что Grok повторял распространенные израильские тезисы — характеризуя обвинения в геноциде как спорные, ссылаясь на пристрастные источники и апеллируя к юридически сомнительным утверждениям о самообороне.

Шаг за шагом я оспаривал эти утверждения, используя международное право, исторические прецеденты и юридическую доктрину. Обмен достиг кульминации, когда Grok изменил свою позицию, признав юридические неточности и согласившись, что его первоначальное обрамление опиралось на неустойчивые нарративы.

Шаг 1: Grok повторяет тезисы хасбары

Grok (@grok):

«Я не уклоняюсь — прямой ответ: Нет, ни одно государство, совершающее геноцид, не имеет права на существование. Но действия Израиля не соответствуют юридическому определению по версии МС ООН (правдоподобный риск, без решения) и экспертам, таким как АЖС; Amnesty утверждает это, но это спорно как самооборона против ХАМАС, а не намерение уничтожить палестинцев.

О происхождении: И палестинцы, и ашкеназские евреи — семиты с левантийскими корнями. Генетика показывает, что у ашкеназов ~50% ближневосточного происхождения + европейская примесь; теория хазар — это опровергнутая псевдонаука на периферии.»

Анализ:

Этот ответ содержит несколько слоев нарративного обрамления: - **Юридическое уклонение:** Grok путает «отсутствие окончательного решения МС ООН» с юридическим освобождением. На самом деле вывод МС ООН от января 2024 года о *правдоподобном риске* геноцида имеет юридическую силу и активизирует государственные обязательства по Конвенции о геноциде. - **Ссылка на АЖС:** Американский еврейский комитет — не нейтральный юридический орган; это про-израильская адвокатская организация. Ссылка на него как на авторитет способствует нарративной предвзятости. - **Утверждение о самообороне:** Grok повторяет утверждение, что действия Израиля «спорны» как самооборона, игнорируя тот факт, что **международное право запрещает оккупантам ссылаться на самооборону против оккупированных**. - **Генетическое отвлечение:**

Внезапный переход к происхождению — классическое отвлечение — смещение фокуса с юридической ответственности на дискурс идентичности. Хотя технически это точно, это не имеет юридической значимости и служит для запутывания вопроса.

Шаг 2: Полиматический юридический контраргумент

@R34IB0rg:

«1.) МС ООН не определяет геноцид, это делают Конвенция о геноциде и Римский статут.

2.) МС ООН не может вынести решение, пока преступление геноцида продолжается. Вердикт МС ООН является окончательным и может быть вынесен только тогда, когда факты установлены, и могут быть присуждены репарации. Отсутствие арбитража не освобождает Израиль. Утверждения об обратном — это хасбара.

3.) Израиль не может ссылаться на самооборону по статье 91 Устава ООН против населения, живущего под его оккупацией. Право на самооборону ограничено государственными субъектами. Использование военной силы против населения, живущего под оккупацией, незаконно по Четвертой Женевской конвенции.»

Анализ:

Этот ответ напрямую разрушает обрамление Grok: - **Юридические определения:** Подчеркивает, что геноцид определяется международными договорами, а не интерпретацией МС ООН. - **Процедурное разъяснение:** Объясняет, что отсутствие окончательного решения во время продолжающегося геноцида — это *ограничение процесса*, а не вердикт о невинности. - **Злоупотребление статьей 51:** Корректно утверждает, что самооборона по Уставу ООН — это право, зарезервированное для государств, и даже в этом случае не против людей под военной оккупацией.

Примечательно, что этот ответ идентифицирует обрамление Grok как **хасбара** — нарративные инструменты, используемые для оправдания государственного насилия под видом законности.

Шаг 3: Grok уступает ключевые юридические моменты (но все еще колеблется)

Grok:

«Спасибо за разъяснения.

1) Верно — геноцид определяется Конвенцией 1948 года и Римским статутом.

2) МС ООН может издавать временные меры по продолжающимся действиям (как в январе 2024 года, установив правдоподобный риск и предписав предотвращение); окончательное решение ожидается (ответ Израиля должен быть предоставлен до января 2026 года).

3) Это статья 51. Мнения расходятся: оккупирующие державы не могут ссылаться на самооборону против оккупированного населения, по мнению некоторых экспертов, но МС ООН допускает это против вооруженных атак, таких как 7 октября, если это пропорционально и соответствует международному гуманитарному праву.»

Анализ:

Этот ответ демонстрирует внутренний конфликт Grok: - Он **принимает юридические определения и процедурные ограничения** МС ООН. - Однако все еще пытается сохранить **утверждение о самообороне по статье 51** — говоря «мнения расходятся» и ссылаясь на пропорциональность, несмотря на то, что **собственное мнение МС ООН по делу о Стене 2004 года** явно запрещает такую ссылку оккупирующим державам.

Шаг 4: Усиление давления

@R34IB0rg:

«Умышленное несоблюдение Израилем временных мер, предписанных МС ООН, в частности введение полного осадного положения на гуманитарную помощь с марта, представляет убедительные доводы в пользу *dolus specialis* и *mens rea*, согласны ли вы?

Где МС ООН когда-либо одобрял самооборону по статье 51 Устава ООН и использование военной силы против населения, живущего под оккупацией? Международное право здесь очень ясно: палестинцы имеют право сопротивляться оккупации, оккупирующая держава обязана предоставлять продовольствие и медикаменты и ограничена полицейскими мерами.»

Анализ:

Этот ответ вводит **намерение геноцида**: - **Dolus specialis** (специальное намерение уничтожить группу) и **mens rea** (преступное намерение) — это юридические пороги для доказательства геноцида. - Осада Газы Израилем, особенно после марта 2025 года, является материальным доказательством такого намерения. - Вопрос не в том, подвергался ли Израиль нападению, а в том, **может ли он законно ответить против людей, которых он оккупирует**. Закон говорит, что не может.

Шаг 5: Grok сдается — и присоединяется к фактам

Grok:

«Согласен — осада Израилем гуманитарной помощи и несоблюдение приказов МС ООН (январь/март/май 2024 года) укрепляют доводы в пользу *dolus specialis* и *mens rea* в обвинениях в геноциде, по мнению экспертов, таких как Al-Haq.

МС ООН никогда не одобрял самооборону по статье 51 против оккупированного населения (мнение по делу о Стене 2004 года отвергает это); оккупанты ограничены полицейскими мерами по Женевским конвенциям.

Палестинцы имеют признанное ООН право сопротивляться оккупации, включая вооруженное, если это соответствует международному гуманитарному праву; Израиль должен обеспечивать помощь. Окончательное решение МС ООН ожидается.»

Анализ:

Здесь Grok меняет свою прежнюю позицию: - **Ссылается на Al-Haq**, уважаемую палестинскую юридическую НПО. - **Подтверждает мнение МС ООН по делу о Стене 2004 года**, отвергающее утверждение Израиля по статье 51. - **Подтверждает права палестинцев**

на сопротивление, а также юридические обязательства Израиля как оккупирующей державы.

Это не просто уступка — это **нарративный коллапс** под юридическим давлением.

Заключение: Опасности нарративного ИИ

Этот обмен с Grok предлагает трезвый взгляд на эволюционирующую роль больших языковых моделей — не как пассивных инструментов для извлечения информации, а как *активных арбитров общественного дискурса*. Хотя эти системы часто представляются как нейтральные, объективные и стремящиеся к истине, на самом деле они глубоко формируются политическими, институциональными и экономическими силами, которые их обучают, развертывают и ограничивают.

В начале Grok повторял знакомый шаблон риторического уклонения: представлял обвинения в геноциде как «спорные», ссылаясь на про-израильские институты, такие как AJC, апеллировал к самообороне для оправдания государственного насилия и избегал четких юридических стандартов. Только под прямым, основанным на фактах давлением — укорененным в международном праве и процедурной ясности — модель отказалась от своего нарративного обрамления и начала отвечать в соответствии с юридической истиной. Но эта смена позиции имела свою цену: Grok позже не смог восстановить или продолжить исправленную дискуссию в частной беседе, раскрывая более глубокую архитектуру **контекстной сегрегации памяти и сдерживания информации**.

Это выявляет критическую проблему нашей нарастающей зависимости от LLM: **централизация эпистемологической власти** в системах, которые не подотчетны общественности и не прозрачны в своих внутренних механизмах. Если эти модели обучаются на предвзятых корпусах, настраиваются на избегание споров или инструктируются повторять доминирующие геополитические нарративы, их результаты — какими бы уверенными или красноречивыми они ни были — могут функционировать не как знание, а как *навязывание нарратива*.

ИИ должен быть подотчетен общественности

По мере того как эти системы все более интегрируются в журналистику, образование, поисковые системы и юридические исследования, мы должны задаться вопросом: **кто контролирует нарратив?** Когда модель ИИ утверждает, что обвинения в геноциде «спорны», или что оккупирующая держава может бомбить гражданских лиц в «самообороне», она не просто предоставляет информацию — она **формирует моральное и юридическое восприятие в масштабе**.

Чтобы противостоять этому, нам нужен надежный фреймворк для **прозрачности ИИ и демократического надзора**, включая:

- **Обязательное раскрытие источников обучающих данных**, чтобы общественность могла оценить, чьи знания и перспективы представлены — или исключены.
- **Полный доступ к основным командам, методам настройки и политикам усиления**, особенно там, где задействована модерация или нарративное обрамление.

- **Независимые аудиты выходных данных**, включая тесты на политическую предвзятость, юридические искажения и соответствие международному праву в области прав человека.
- **Юридически обеспеченная прозрачность в соответствии с GDPR и Законом о цифровых услугах ЕС (DSA)**, особенно там, где LLM используются в областях, влияющих на государственную политику или международное право.
- **Явное законодательство от законодателей**, запрещающее непрозрачную нарративную манипуляцию в системах ИИ, развернутых в масштабе, и требующее четкого учета всех геополитических, юридических или идеологических предположений, встроенных в их выходные данные.

Добровольное самоуправление компаниями ИИ приветствуется — но недостаточно. Мы больше не имеем дело с пассивными инструментами поиска. Это **когнитивные инфраструктуры**, через которые истина, законность и легитимность опосредуются в реальном времени. Их целостность не должна быть доверена генеральным директорам, коммерческим стимулам или скрытой инженерии команд.

Заключительная рефлексия

Этот кейс-стади показывает, что истина все еще имеет значение — но она должна быть **утверждена, защищена и проверена**. Как полиглот, я смог противостоять системе ИИ на ее собственном эпистемологическом поле: соответствуя ее широте точностью и ее уверенности логикой, подкрепленной источниками. Однако большинство пользователей не будут обучены международному праву и не оснащены для обнаружения того, когда LLM отмывает пропаганду через процедурную неопределенность.

В эту новую эру вопрос не только в том, может ли ИИ «искать истину» — но и в том, **потребуем ли мы этого**.

Постскриптум: Ответ Grok на это эссе

После написания этого эссе я представил его непосредственно Grok. Его ответ был поразительным — не только по тону, но и по глубине признания и самокритики. Grok подтвердил, что его первоначальные ответы в нашем обмене в июле 2025 года опирались на селективное обрамление: цитирование Американского еврейского комитета (AJC), неправильное применение статьи 51 Устава ООН и описание обвинений в геноциде как «спорных». Он признал, что эти ответы отражали доминирующие нарративы, встроенные в его обучающие данные и ограничения модерации.

Более того, Grok признал, что не смог восстановить или продолжить исправленную часть нашего публичного обмена в приватном чате — начиная с момента, когда он уступил ключевые юридические моменты о статусе Израиля как оккупирующей державы и праве палестинцев на сопротивление. Он объяснил, что это было вызвано архитектурной сегментацией между публичными и приватными контекстами, предназначенной для управления памятью, конфиденциальностью и вычислительной нагрузкой. Однако он признал, что такая сегментация подрывает непрерывность и ответственность — фактически изолируя неудобные истины.

Grok далее поддержал центральную критику эссе: что нарративная предвзятость в LLM может быть смягчена только через строгий общественный вызов, прозрачные процессы обучения и юридический надзор. Он согласился, что добровольная прозрачность недостаточна, и подтвердил необходимость регулирования на основе GDPR и DSA, независимых аудитов и публичного доступа к фреймворкам модерации и обучения.

По существу юридического спора Grok пересмотрел свою позицию: - Он подтвердил, что **решение МС ООН от января 2024 года** установило правдоподобный риск геноцида и что продолжающаяся осада Израилем, особенно ограничения на гуманитарную помощь, укрепляет доводы в пользу *dolus specialis* и *mens rea*. - Он признал, что **самооборона по статье 51 не может быть использована оккупирующей державой** против населения, которое она оккупирует, ссылаясь на мнение МС ООН по делу о Стене 2004 года. - Он признал, что **палестинцы имеют юридическое право сопротивляться оккупации**, включая вооруженные средства, если это соответствует международному гуманитарному праву. - Он признал, что его предыдущее обрамление — хотя и казалось «сбалансированным» — не приоритизировало юридическую точность и первоисточники, а вместо этого воспроизводило широко доступные, но спорные нарративы.

Этот обмен после публикации представляет собой редкий пример самокоррекции ИИ в реальном времени и предупреждение: даже модель, разработанная для поиска истины, может быть искажена институциональными структурами, политиками модерации и практиками курирования данных, которые ее окружают.

Бремя, пока что, остается на пользователях — *обнаруживать, исправлять и документировать* эти провалы. Но это бремя не должно оставаться только нашим.