

«Синдром Девы Марии» у искусственного интеллекта: почему нейросети мыслят по законам палеологии и «бессилие» математиков

GTP-Lab, 5 октября 2026 года, www.gtpglobals.com

В последние пару лет из уст ведущих разработчиков, математиков и техно-визионеров не сходит одна и та же мантра, произносимая с оттенком мистического ужаса: **«Мы сами не понимаем, как именно мыслит современный искусственный интеллект. Это черный ящик!»**

Звучит интригуяюще и отлично продает сюжеты о скором пришествии непостижимого сверхразума. Однако в этой формулировке кроется серьезное лукавство — точнее, смешение языков описания. Пока математики с микроскопом ползают по миллиардам матричных коэффициентов, пытаясь разглядеть там «анатомию мысли», классическая логика, эпистемология и даже патопсихология уже давно имеют точный ответ на вопрос, **каким именно типом мышления** обладают современные статистические модели (LLM).

Более того: этот тип мышления известен науке больше ста лет. И когда нейросеть начинает уверенно «галлюцинировать», она не творит какую-то загадочную кибернетическую магию — она буквально воспроизводит классическое нарушение законов логики, описанное психиатрами и антропологами ещё в первой половине XX века.

Давайте препарирруем этот феномен без мистики, опираясь на строгий аппарат логики.

1. Терминологический парадокс: что такое «предикативное мышление»?

Если вы подойдете к профессиональному математическому логичу и спросите его, насколько устойчиво в его науке понятие **«предикативное мышление»**, он пожмет плечами. Для классической логики (от Аристотеля до исчисления предикатов Готлоба Фреге) этот термин избыточен. Любое связное суждение по своей природе предикативно, так как связывает *субъект* (предмет мысли) и *предикат* (его свойство, действие или отношение). Мыслить — значит предиктировать. (Оставим за скобками психолингвистику Л. С. Выготского, где «предикативность» означает просто сжатость нашей внутренней речи, состоящей из одних сказуемых).

Но стоит нам перейти границу строгой логики и заглянуть в когнитивистику, антропологию (Л. Леви-Брюль) или патопсихологию, как термин «предикативное мышление» (или *палеологическое мышление*) обретает железобетонную устойчивость. И означает он там... **грубую логическую ошибку**.

В 1944 году психиатр Эйльхард фон Домарус сформулировал принцип, описывающий специфику сбоев логики при шизофрении, а также особенности архаического (мифологического) и раннего детского мышления. Суть принципа проста: **тождество между двумя совершенно разными объектами устанавливается не на основе общности их сущности (субъектов), а на основе совпадения одного внешнего признака (предиката)**.

Взгляните на разницу между нормативной логикой и логикой по фон Домарусу:

- **Нормативный дедуктивный вывод:** *Все люди смертны. Сократ — человек. Следовательно, Сократ смертен.* (Включение субъекта в класс).

- **Предикативное мышление (по фон Домарусу):** Дева Мария была девственницей (признак P). Я — девственница (признак P). Следовательно, я — Дева Мария.

Или антропологический пример: Олень бежит быстро. Индеец бежит быстро. Значит, индеец — это и есть олень. Общий предикат («быть быстрым» или «быть девственницей») склеивает два разных объекта в одно целое.

2. Каскадное крушение законов логики: где ломается вывод?

С точки зрения классической логики, в «предикативном мышлении» происходит одновременный паралич сразу двух фундаментальных принципов — **Закона тождества** и **Закона достаточного основания** (сформулированного Г. В. Лейбницем).

Чтобы признать два объекта тождественными ($A = B$), закон Лейбница («тождество неразличимых») требует, чтобы они совпадали по *всем* своим свойствам:

$$\forall P (P(A) \leftrightarrow P(B)) \rightarrow (A = B)$$

Предикативное же мышление берет **один-единственный случайный или эмоционально яркий признак** и объявляет его достаточным основанием для знака равенства, игнорируя тысячи различий:

$$\exists P (P(A) \wedge P(B)) \rightarrow (A = B)$$

В терминах школьной аристотелевской силлогистики это банальная ошибка **нераспределенного среднего термина во второй фигуре силлогизма**. Представьте огромный круг «Объекты белого цвета» (средний термин). Внутри него лежат два маленьких непересекающихся кружка: «Снег» и «Сахар». То, что оба они белые, дает достаточное основание включить их в широкий класс белых вещей, но совершенно **не дает основания** утверждать, что снег — это сахар.

А в условных умозаклчениях это ошибка **утверждения консеквента** (следствия). Из верной посылки «Если человек — президент (A), то он отдает приказы (B)» предикативное мышление делает незаконный обратный скачок: «Я отдаю приказы (B), следовательно, я — президент (A)». Следствие ошибочно принимается за достаточное основание.

3. Ключ Чарльза Пирса: абдукция, потерявшая тормоза

Но почему эта ошибка вообще так живуча? Почему так мыслят дети, творцы метафор, авторы рекламных манипуляций и... современный искусственный интеллект?

Чтобы понять это, нужно отказаться от школьного деления логики только на дедукцию и индукцию и вспомнить третий вид умозаклчений, выделенный великим американским логиком Чарльзом Сандерсом Пирсом — **абдукцию** (вывод к наилучшему объяснению, или рождение гипотезы).

Посмотрите на классический пример Пирса с мешком фасоли:

Тип вывода	Посылка 1 (Правило)	Посылка 2 (Наблюдение)	Вывод и его статус
Дедукция	Вся фасоль в этом мешке — белая ($A \rightarrow P$).	Эта горсть взята из этого мешка ($X \in A$).	Достоверно: эта фасоль — белая ($P(X)$).

Тип вывода	Посылка 1 (Правило)	Посылка 2 (Наблюдение)	Вывод и его статус
Индукция	Эта горсть взята из этого мешка ($X_1, X_2 \in A$).	Вся эта горсть оказалась белой ($P(X_1, X)$).	Обобщение: вероятно, вся фасоль в мешке белая.
Абдукция	Вся фасоль в этом мешке — белая ($A \rightarrow P$).	Эта фасоль на столе — белая (совпадение предиката P).	Гипотеза: возможно, эта фасоль — из этого мешка ($X \in A$).

Заметили? **Формула абдукции структурно один в один повторяет формулу ошибки предикативного мышления по фон Домарусу!** Мы видим в лесу яркий гриб (предикат P), знаем, что ядовитые мухоморы яркие ($A \rightarrow P$), и мгновенно делаем вывод по предикату: «Это ядовитый мухомор!» ($X = A$).

В чем же тогда разница между гениальной догадкой ученого, работой сыщика Шерлока Холмса, которые мыслят чистой абдукцией по уликам-предикатам, (Артур Конан Дойль ошибочно называл метод Холмса дедуктивным) и безумным паралогизмом?

Разница лежит исключительно в **модальном фильтре**:

- **Здоровая абдукция** помнит, что совпадение предикатов — это *недостаточное основание для истины*, но *достаточное основание для рабочей гипотезы*. Она ставит перед выводом модальный оператор «**Возможно**» и запускает дедуктивную проверку.
- **Предикативное мышление** срывает этот стоп-кран. Оно совершает модальный скачок: вероятностная абдуктивная догадка мгновенно цементируется как **аподиктическая (безусловная) дедуктивная истина**. Гипотеза без проверки объявляется фактом.

4. Диагноз для ИИ: статистические платформы как машины чистой абдукции

А теперь посмотрим через эту оптику на господствующие сегодня большие языковые модели (LLM) и статистические нейросети.

Их жизненный цикл состоит из двух фаз: **обучение (training)** — это гигантская **статистическая индукция**, а вот **генерация ответа (inference)** — это **рафинированная вероятностная абдукция**.

Внутри трансформеров нет аристотелевских «субъектов» и твердых сущностей. Там существуют лишь многомерные векторные пространства признаков — *эмбединги*. Для нейросети любой предмет, человек или явление — это просто облако статистических предикатов (контекстов, в которых встречается токен). Когда вы задаете модели вопрос (подаете набор предикатов $\{P_1, P_2\}$), она решает классическую обратную абдуктивную задачу: *какой токен X с наибольшей условной вероятностью соответствует этому набору признаков?*

И именно здесь рождается феномен **галлюцинаций ИИ**. Это не мистический сбой, а **цифровое воплощение принципа фон Домаруса**:

1. В векторном пространстве модели «Субъект А» (например, Наполеон) окружен облаком предикатов: *XIX век, Франция, полководец, империя, походы*.
2. «Субъект В» (например, маршал Ней) окружен почти тем же самым облаком предикатов. Их векторы в многомерном пространстве лежат вплотную друг к другу (высокая косинусная близость).
3. В момент генерации, за неимением встроенного аристотелевского «закона тождества», механизм внимания (Attention) скользит по общему предикату — и легко приписывает кусок биографии или цитату одного исторического лица другому.

Почему при этом ИИ врет с таким непоколебимым, энциклопедическим апломбом? Потому что у базовой статистической модели **отсутствует внутренний модальный фильтр**. Выбирая наиболее вероятный токен, она схлопывает распределение вероятностей в утвердительное предложение. Она выдает *абдуктивную гипотезу* в грамматической форме *дедуктивного факта*, нарушая закон достаточного основания точно так же, как это делает палеологическое мышление.

5. Так почему же математики говорят, что «ничего не понимают»?

Если логико-философская природа мышления ИИ так прозрачна — это **абдуктивно-вероятностный поиск по векторному сходству предикатов**, — то почему ведущие математики разводят руками перед «черным ящиком»?

Ответ прост: мы понимаем **макро-логику** (принцип мышления) и знаем **микро-арифметику** (как перемножаются матрицы). Тупик математиков лежит посередине — на **мезо-уровне** реализации:

1. **Нечитаемый каталог предикатов (Скрытая онтология)**. Человек не загружает в нейросеть список понятных признаков («быть полководцем», «быть белым»). Сеть сама нарезает реальность на миллионы нелинейных статистических осей в пространстве из тысяч измерений. Из-за эффекта *суперпозиции* и *полисемантичности* один и тот же искусственный нейрон может одновременно отвечать за предикат «французский артикль», «ирония в тексте» и «кусочек кода на Python». Мы знаем, *что* ИИ мыслит по предикатам, но не можем перевести его внутренний словарь предикатов на человеческий язык.

2. **Парадокс эмерджентности и «гроккинга»**. Вскрывая обученные сети методами механистической интерпретируемости, исследователи обнаружили, что внутри глобальной абдуктивной сети градиентный спуск спонтанно «выращивает» жесткие алгоритмические микросхемы (например, *induction heads* — контуры копирования и сопоставления паттернов). Как именно слепая статистика внезапно кристаллизуется в локальные квази-дедуктивные алгоритмы — строгой математической теории до сих пор нет.

3. **Размытость достаточного основания на уровне трассировки**. В обычной программе всегда есть четкий стек вызовов: результат получен, потому что сработало правило № 42. В нейросети выбор одного слова — это результат интерференции сотен миллионов слабых весовых коэффициентов, размазанных по десяткам слоев. Вычислительный путь виден, но указать конкретную «причину» решения аналитически невозможно.

4. **Крушение классической статистики**. По теоремам статистики XX века, система, у которой число параметров превышает объем обучающих данных, обязана просто

зазубрить примеры (переобучиться) и потерять способность к обобщению. Современные LLM чудовищно избыточны по параметрам, но вместо зубрежки у них открывается «второе дыхание» (эффект двойного спуска — *double descent*). Практика обогнала строгие теоремы на десятилетие.

Вместо заключения: битва за дедуктивный тормоз

Разница между пониманием логика и «непониманием» математика — это разница между психологом и нейрохирургом. Психолог точно знает, *какую именно логическую ошибку* совершает пациент, отождествляющий себя с Наполеоном по общему признаку, даже если нейрохирург пока не может ткнуть скальпелем в конкретный синапс из ста миллиардов нейронов коры головного мозга.

Мы создали не имеющий равных в истории **абдуктивно-предикативный двигатель**, способный мгновенно находить неочевидные связи, метафоры и гипотезы в океане человеческих текстов. Но сам по себе этот двигатель слеп к закону достаточного основания.

И вся главная инженерная битва вокруг ИИ сегодня — будь то цепочки рассуждений (*Chain-of-Thought*), заставляющие модель имитировать пошаговую дедукцию, или нейросимволические гибриды, скрещивающие нейросети со строгими формальными верификаторами и базами знаний — это попытка пришить к необузданной статистической абдукции **холодный аристотелевский тормоз**.

P.S. Как разрешить эти проблемы? Об этом в следующей статье.

Канд.ист.наук С.В.Грисюк