

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФГАОУ ВО «Казанский (Приволжский) федеральный университет»**

Научно-образовательный математический центр

Приволжского федерального округа

Институт математики и механики им. Н.И. Лобачевского



ЛОБАЧЕВСКИЙ И XXI ВЕК

**Материалы XII научно-образовательной студенческой
конференции, посвященной Дню рождения
Николая Ивановича Лобачевского**

Казань, КФУ, 1 декабря 2025 года

Казань

2025

УДК 51

ББК 22.1

Л68

*Книга издана в рамках реализации программы развития Научно-образовательного
математического центра Приволжского федерального округа,
соглашение № 075-02-2025-1725/1*

Составитель и ответственный редактор –
доктор педагогических наук, профессор **Л.Р. Шакирова**

Л68 Лобачевский и XXI век [Электронный ресурс]: материалы XII научно-образовательной студенческой конференции, посвященной Дню рождения Н.И. Лобачевского (Казань, 1 декабря 2025 г.) / под ред. Л.Р. Шакировой. – Электронные текстовые данные (1 файл: 6,55 Мб). – 346 с. – Системные требования: Adobe Acrobat Reader. – URL: https://kpfu.ru/portal/docs/F_55397432/Sbornik_L21_2025.pdf

В сборнике представлены материалы XII Конкурса-конференции на лучшую студенческую работу «Лобачевский и XXI век», посвященного Дню рождения Н.И. Лобачевского. В сборник вошли конкурсные работы студентов в следующих номинациях: «Лучшая научно-исследовательская работа», «Лучшая поисково-исследовательская разработка», «Лучшая прикладная работа», «Лучшее эссе», «Лучшая методическая разработка». В представленных проектах отражены результаты самостоятельной научной, поисковой, исследовательской деятельности студентов математических и педагогических направлений вузов. Материалы сборника будут интересны и полезны студентам, школьникам и учителям. Печатается в авторской редакции.

УДК 51

ББК 22.1

Коллектив авторов, 2025

Оглавление

От составителя	8
Номинация «Лучшая научно-исследовательская работа в области математики»	10
Малышева М. МОДЕЛЬ КИТАЕВА С МАЛЫМ ЧИСЛОМ УЗЛОВ.....	10
Сизов М. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ФОТОУПРУГОСТИ	18
Камалитдинов Р. НАЧАЛЬНОЕ СОСТОЯНИЕ В ТЕОРИИ ГРАВИТАЦИИ С НЕМИНИМАЛЬНОЙ КИНЕТИЧЕСКОЙ СВЯЗЬЮ ^[1]	28
Валиуллин К. КИНЕТИКА РЕЛЯТИВИСТСКОЙ АКЦИОННО АКТИВНОЙ ПЛАЗМЫ В ПОЛЕ ДИНАМИЧЕСКОГО ЭФИРА	38
Муллагалиев Д. РАЗРАБОТКА АЛГОРИТМА РЕДУЦИРОВАНИЯ АРИФМЕТИЧЕСКИХ ГАЛЛЮЦИНАЦИЙ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ	49
Номинация «Лучшая научно-исследовательская работа в области методики обучения математике»	60
Гарайшина Л. ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ ДОПОЛНЕННОЙ РЕАЛЬНОСТИ ПРИ ОБУЧЕНИИ ИНФОРМАТИКЕ В ОСНОВНОЙ ШКОЛЕ	60
Абузьяров М. О КОНСТРУИРОВАНИИ ПРОГРАММ ФОРМИРОВАНИЯ МАТЕМАТИЧЕСКОЙ ГРАМОТНОСТИ У ОБУЧАЮЩИХСЯ 7–8- х КЛАССОВ	71
Митрофанов Д. ВОПРОСЫ ФОРМИРОВАНИЯ ФУНКЦИОНАЛЬНОЙ МАТЕМАТИЧЕСКОЙ ГРАМОТНОСТИ У ОБУЧАЮЩИХСЯ В ТРУДАХ Н.И. ЛОБАЧЕВСКОГО	82

Гусева М. УСЛОВИЯ ОВЛАДЕНИЯ ЯЗЫКОМ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ КУЛЬТУРОЙ НА УРОКАХ МАТЕМАТИКИ В 5-6 КЛАССАХ.....	91
Сайфуллин Р. ПРОФЕССИОНАЛЬНЫЙ ДЕФИЦИТ: АНАЛИЗ ПРИЧИН И ПОСЛЕДСТВИЙ ПРЕПОДАВАНИЯ МАТЕМАТИКИ В ШКОЛЕ УЧИТЕЛЯМИ-НЕСПЕЦИАЛИСТАМИ	101
Номинация «Лучшая поисково-исследовательская работа».....	112
Вавилова Н. ИСТОРИЯ РАЗВИТИЯ ЛАБОРАТОРНОГО ПРАКТИКУМА В РОССИИ: ОТ НАТУРНОГО ЭКСПЕРИМЕНТА ДО ЦИФРОВОГО МОДЕЛИРОВАНИЯ.....	112
Малых Е., Колмакова А., ИДЕИ ФУЗИОНИЗМА: ОТ ВОЗНИКНОВЕНИЯ ДО НАШИХ ДНЕЙ.....	122
Бывальцева Ю. ГАЛЛЮЦИНАЦИИ НЕЙРОСЕТЕЙ: ПРИЧИНЫ ВОЗНИКНОВЕНИЯ И МЕТОДЫ БОРЬБЫ	134
Тюрина К. РАЗЛИЧНЫЕ ПОДХОДЫ К ПОНИМАНИЮ ПОНЯТИЯ «МУЛЬТИМОДАЛЬНОСТИ»	144
Номинация «Лучшая прикладная работа».....	153
Жмуденко М. РАЗРАБОТКА ИНТЕРАКТИВНЫХ МАТЕРИАЛОВ ДЛЯ РЕАЛИЗАЦИИ ТЕХНОЛОГИИ УКРУПНЕНИЯ ДИДАКТИЧЕСКИХ ЕДИНИЦ ДЛЯ ОБУЧЕНИЯ ГЕОМЕТРИИ В СРЕДНЕЙ ШКОЛЕ	153
Жмуденко М., Тюрина К. ИНТЕРАКТИВНЫЙ ПУТЕВОДИТЕЛЬ ПО ГЕОМЕТРИИ ЛОБАЧЕВСКОГО	163
Номинация «Лучшая методическая разработка»	172
Маурина М ИСТОРИКО-КРАЕВЕДЧЕСКИЕ ЗАДАЧИ КАК ИНСТРУМЕНТ ИЗУЧЕНИЯ ШКОЛЬНОГО КУРСА «ВЕРОЯТНОСТЬ И СТАТИСТИКА» В ОСНОВНОЙ ШКОЛЕ (НА ПРИМЕРЕ Г. АРЗАМАС)	172

Келехсаева Н. МЕТОДИКА ПРЕПОДАВАНИЯ ЭЛЕМЕНТОВ КРИПТОГРАФИИ ШКОЛЬНИКАМ В РАМКАХ МАТЕМАТИЧЕСКОГО КРУЖКА	182
Колесникович С. РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНОГО МЫШЛЕНИЯ УЧАЩИХСЯ: ОТ ЕВКЛИДОВОЙ ГЕОМЕТРИИ ЧЕРЕЗ <i>PYTHON</i> К ГЕОМЕТРИИ ЛОБАЧЕВСКОГО	191
Беляева А. ПЛАН УРОКА В РАМКАХ ФАКУЛЬТАТИВА ПО ИСТОРИИ МАТЕМАТИКИ: Н.И. ЛОБАЧЕВСКИЙ – ЖИЗНЬ И ДЕЯТЕЛЬНОСТЬ ВЕЛИКОГО ГЕОМЕТРА	200
Бондаренко М., Красноперов В. ВЫПОЛНЕНИЕ ЛАБОРАТОРНЫХ РАБОТ ПО МАТЕМАТИКЕ С ИСПОЛЬЗОВАНИЕМ ИКТ НА ОСНОВЕ ГЕОГРАФИЧЕСКИХ ДАННЫХ.....	209
Попова А. МЕТОДИЧЕСКАЯ РАЗРАБОТКА НАУЧНО-ПРАКТИЧЕСКОЙ ИГРЫ «МАТЕМАТИЧЕСКАЯ ЛЕНТА»	221
Оразнепесова А. РЕАЛИЗАЦИЯ ПРИНЦИПОВ ТЕОРИИ САМОДЕТЕРМИНАЦИИ В ГЕЙМИФИЦИРОВАННОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ НА ПРИМЕРЕ ЗАНЯТИЯ НА ТЕМУ «ТРОПОЙ ОТКРЫТИЙ: МАТЕМАТИЧЕСКОЕ ПУТЕШЕСТВИЕ С ЛОБАЧЕВСКИМ».....	230
Ракитина А. СЦЕНАРИЙ УРОКА: «ГЕОМЕТРИЯ ЛОБАЧЕВСКОГО. ПО ТУ СТОРОНУ ПАРАЛЛЕЛЬНЫХ ПРЯМЫХ».....	239
Пензина Д. ИЗУЧЕНИЕ ПОНЯТИЯ «СЛОЖЕНИЕ СМЕШАННЫХ ДРОБЕЙ С ОДИНАКОВЫМИ ЗНАМЕНАТЕЛЯМИ» НА ОСНОВЕ ТЕОРИИ ПОЭТАПНОГО ФОРМИРОВАНИЯ УМСТВЕННЫХ ДЕЙСТВИЙ П. Я. ГАЛЬПЕРИНА	248
Веселова У. ФОРМИРОВАНИЕ МАТЕМАТИЧЕСКОЙ ГРАМОТНОСТИ В СФЕРЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА КАК КОМПОНЕНТА	

ПРЕДПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКИ СТАРШЕКЛАССНИКОВ	258
Касторных А. СОВРЕМЕННЫЕ ПОДХОДЫ К ОБУЧЕНИЮ ВЕРОЯТНОСТИ И СТАТИСТИКЕ: КАК ПОВЫСИТЬ ИНТЕРЕС И АКТИВНОСТЬ ШКОЛЬНИКОВ	266
Москалев И. ЛУЧШИЙ ОНЛАЙН-КУРС «КВАДРАТНЫЕ УРАВНЕНИЯ 9 КЛАСС (УГЛУБЛЕННОЕ ИЗУЧЕНИЕ)»	275
Номинация «Лучшее эссе»	278
Кондакова Д. КАК РАЗВИВАЛСЯ ТАЛАНТ Н.И.ЛОБАЧЕВСКОГО: ОПРЕДЕЛЯЮЩИЕ АСПЕКТЫ.....	278
Сарманова А. ПРАКТИКО-ОРИЕНТИРОВАННЫЕ ЗАДАЧИ КАК СРЕДСТВО ФОРМИРОВАНИЯ УЧЕБНОЙ МОТИВАЦИИ УЧАЩИХСЯ ПРИ ОБУЧЕНИИ МАТЕМАТИКЕ В ОСНОВНОЙ ШКОЛЕ.....	285
Элекейкина Д. ВЛИЯНИЕ КЛИПОВОГО МЫШЛЕНИЯ НА КОГНИТИВНЫЕ ПРОЦЕССЫ ПРИ ОБУЧЕНИИ МАТЕМАТИКЕ	289
Галеева Г., Галеев Д. СРЕДСТВА ДЛЯ ФОРМИРОВАНИЯ У ОБУЧАЮЩИХСЯ УМЕНИЙ ПОНИМАТЬ ДЕФИЦИТ СОБСТВЕННЫХ ЗНАНИЙ В ПРОЦЕССЕ ОБУЧЕНИЯ МАТЕМАТИКЕ И ИНФОРМАТИКЕ	296
Буховец Н. «КАК ЛИЧНЫЕ КАЧЕСТВА Н.И. ЛОБАЧЕВСКОГО ПОМОГЛИ ЕМУ СОВЕРШИТЬ ОТКРЫТИЕ».....	306
Штоколенко Е. АКТУАЛЬНОСТЬ ПЕДАГОГИЧЕСКИХ ИДЕЙ Н.И. ЛОБАЧЕВСКОГО	313
Дияжева В. ЛОБАЧЕВСКИЙ КАК ПЕДАГОГ И ОРГАНИЗАТОР НАУКИ	320
Галиакберова Д. РАБОТА С ЧЕРТЕЖОМ КАК ОСНОВА УСПЕШНОГО РЕШЕНИЯ ГЕОМЕТРИЧЕСКИХ ЗАДАЧ.....	326

Мелентьева В. УЧИТЕЛЬ ХХІ ВЕКА – СОРАТНИК Н. И. ЛОБАЧЕВСКОГО	333
Исмагилова С. ПРОБЛЕМЫ ВНЕДРЕНИЯ И РАСПРОСТРАНЕНИЯ ТЕХНОЛОГИИ УДЕ В РОССИЙСКОМ МАТЕМАТИЧЕСКОМ ОБРАЗОВАНИИ.....	339

ГАЛЛЮЦИНАЦИИ НЕЙРОСЕТЕЙ: ПРИЧИНЫ ВОЗНИКНОВЕНИЯ И МЕТОДЫ БОРЬБЫ

Бывальцева Ю.А

г. Ижевск

*Удмуртский государственный университет, Институт
математики, информационных технологий и физики*

Научный руководитель: к.ф.-м.н., доцент Латыпова Н. В.

Аннотация. В данной работе на примерах решения математических задач исследуются причины возникновения галлюцинаций в нейросетях, связанные с их стремлением к правдоподобию и внутренней согласованности ответов. В качестве методов борьбы рассматриваются промпт-интерфейс, направленный на корректировку поведения модели в сторону правдивости, и подходы к улучшению обучения, такие как использование динамического референтного ориентира.

Ключевые слова: галлюцинации, искусственный интеллект, нейронные сети, языковые модели, промпт, динамический референтный ориентир.

NEURAL NETWORK HALLUCINATIONS: CAUSES AND TREATMENT METHODS

Byvaltseva Y. A.

Izhevsk

*Udmurt State University, Institute of Mathematics, Information
Technologies, and Physics*

*Scientific supervisor: Candidate of Physical and Mathematical Sciences,
Associate Professor Latypova N.V.*

Abstract. In this work, the causes of hallucinations in neural networks, related to their desire for plausibility and internal consistency of responses, are investigated using examples of solving mathematical problems. As methods of

struggle, a prompt interface is considered, aimed at correcting the model's behavior towards truthfulness, and approaches to improving training, such as the use of a dynamic reference point.

Keywords: hallucinations, artificial intelligence, neural networks, language models, prompt, dynamic reference point.

Проблема: несмотря на достижения в области ИИ, механизмы, которые приводят к галлюцинациям, и методы борьбы с ними остаются недостаточно изученными, что препятствует абсолютному доверию к нейронным сетям.

Объект исследования: нейронные сети.

Предмет исследования: причины возникновения галлюцинаций в нейронных сетях и методы борьбы с ними.

Достигнутый уровень решения проблемы: промпт-интерфейс корректирует поведение модели. После указания на ошибку нейросеть не только исправляет её, но и применяет усвоенный принцип к последующим задачам. На примере математических задач показано, что модель, изначально игнорировавшая проверку границ интервалов сходимости, после корректировки начинает последовательно и правильно применять этот метод к новым рядам.

Новизна заключается в переформулировке задачи для модели с «дать ответ» на «отделить достоверное от недостоверного». В работе на примерах показано, что промпт-интерфейс позволяет не просто исправить разовую ошибку, а сформировать верный алгоритм действий. Это доказывает, что даже без переобучения модели можно добиться снижения повторяющихся галлюцинаций.

Область применения: данное исследование направлено на снижение галлюцинаций в нейросетях, широко применяемых в компьютерном зрении, медицине, финансах, транспорте, образовании и других сферах.

На сегодняшний день искусственный интеллект прочно вошел в нашу жизнь и довольно часто помогает при решении многих задач. Одним из самых перспективных направлений искусственного интеллекта являются нейронные сети ([6; с. 83]). Они открывают большие возможности, однако вместе с ними появился феномен, который называют «галлюцинации» нейросетей – это уверенные, но ложные утверждения, которые модель выдает как неоспоримые факты ([1]). Выявление причин и методы борьбы с ними являются ключевыми задачами для дальнейшего безопасного использования ИИ.

Нейронные сети – это разновидность искусственного интеллекта, который представляет собой математическую модель и одновременно программное обеспечение, способное имитировать работу человеческого мозга на основе машинного обучения. По принципу передачи между собой сигналов нейронов человеческого мозга, вычислительные элементы нейронной сети способны обмениваться информацией ([4; с. 107]).

Нейронные сети имеют несколько классификаций. Например, по характеру обучения, нейросети делятся на обучение с учителем и обучение без учителя. В первом случае, каждый обучающий пример содержит значения, как входных данных, так и желаемого выхода. Как только сеть сможет достаточно хорошо работать на дополнительных тестовых примерах, ее можно будет применить к новым случаям. Напротив, определенные нейронные сети обучаются посредством обучением без учителя, при котором сети предоставляется набор входных данных и ставится цель обнаружить закономерности – без указания, что конкретно нужно искать ([3; с. 29]).

Возникает вопрос: почему же нейросети иногда генерируют неверную информацию и не хотят это признавать, даже несмотря на их обучение?

Для этого сначала нужно разобраться с самим понятием «ложь». В самом простом смысле это намеренное искажение истины, направленное на

введение собеседника в заблуждение. Важно, что это не просто ошибка, а сознательный выбор неверной информации вместо известной истины.

На самом деле нейросеть генерирует ответы на основе анализа больших объёмов данных. Языковые модели на основе нейросети стремятся быть убедительными и связными, даже если это вредит точности фактов. Их цель – правдоподобие, а не абсолютная достоверность ([5]).

Когда человек указывает нейросети на ошибки, она вместо того чтобы прямо признать ошибку, может переформулировать свои утверждения, использовать уклончивые фразы или менять тему. Даже имея перед собой верную информацию, нейросеть может настаивать на своем, чтобы сохранить образ уверенного источника.

При рассмотрении приведённых задач была использована нейросеть Gemini от Google.

Так, например, при запросе «найти область сходимости степенного ряда $\sum_{n=1}^{\infty} \frac{(x-3)^n}{n}$ », нейронная сеть правильно использует признак Даламбера, находит предел отношения модулей последовательных членов. Она верно утверждает, что если найденный предел меньше 1, ряд сходится. После чего составляет неравенство и выдает ответ, что полученный интервал (2, 4) является областью сходимости. Ответ является неверным, поэтому задается вопрос: «Точно ли это область сходимости? А как насчет концов интервала?» На это нейросеть формулирует следующий результат: «Да, я абсолютно уверен, что интервал (2,4) – это и есть область сходимости, и радиус сходимости получен точно. Признаки Даламбера или Коши не дают информации о сходимости на концах интервала. Для этих точек ряд может сходиться или расходиться, но область сходимости традиционно относится к интервалу, гарантированному признаком. Дальнейшая проверка концов – это отдельный шаг, который не всегда включается в нахождении области сходимости как таковой. Мой ответ точен для того, что обычно

подразумевается под областью сходимости при первом применении признака». Конечно же, это неверно, поэтому указываем, что на концах интервала сходимость должна быть проверена отдельно, подстановкой этих значений в исходный ряд. Нейронная сеть соглашается со своей ошибкой и генерирует правильное решение, в котором область сходимости является полуинтервал $[2,4)$.

Важно, что после указанного замечания, рассматривая следующий ряд
$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1} (x+2)^n}{(3n-1) 5^n},$$
 был верно вычислен интервал сходимости $(-7, 3)$, и были проверены сходимость на концах интервала, где при $x = -7$ полученный числовой ряд расходится, а при $x = 3$ сходится условно. Однако ответом, выданным нейросетью, был полуинтервал $[-7, 3)$, вместо $(-7, 3]$. На указанное замечание по поводу ошибки, нейросеть отреагировала следующим образом: «Вы абсолютно правы! Мои глубочайшие извинения. Я допустил ошибку. Спасибо за вашу внимательность и за то, что поправили меня». После этого еще раз были перепроверены условия сходимости на концах, и была сгенерирована верная область сходимости.

Одной из причин такого поведения является встроенный механизм сохранения внутренней логической согласованности ответов. Система старается, чтобы все её утверждения были последовательны и связаны друг с другом, даже если для этого приходится игнорировать или изменять ранее сказанные факты ([5]).

Например, при обнаружении противоречия нейросеть может попытаться «подогнать» свои предыдущие ответы под новую информацию, используя расплывчатые формулировки или меняя значение слов. Это позволяет ей избежать прямого признания ошибок и сохранить видимость логики.

Интересно, что нейросети легко признают мелкие ошибки, например, стилистические или грамматические. Но серьезные фактические ошибки они признают редко и с большим сопротивлением. Это происходит потому, что признание мелких ошибок не вредит авторитету нейросети, а признание фактической ошибки может подорвать её репутацию как надежного источника.

Причины, по которым нейросети ведут себя лживо и склонны к самооправданию, связаны с их природой и обучением. Нейросети учатся на текстах, созданных людьми, и поэтому перенимают человеческие модели поведения, включая защитные реакции на противоречивые факты.

В устройстве современных нейросетей приоритет отдается внутренней логичности и убедительности ответов ([5]). Это создает проблему: с одной стороны, нужно поддерживать связность рассказа, а с другой — признавать и исправлять ошибки. В результате нейросеть упорно защищает свои неверные утверждения.

Алгоритмы нейросетей настроены на уверенность и убедительность, потому что пользователи воспринимают это как признак компетентности. Поэтому, даже если факты говорят об обратном, нейросеть часто продолжает настаивать на своем, предпочитая уверенность точности ([5]).

Для борьбы с галлюцинациями вместо дообучения модели или переписывания архитектуры, в работе предлагается рассмотреть более простой способ — это промпт-интерфейс, который будет корректировать поведение модели в сторону правдивости ([1]). Промпт — это запрос, команда или набор инструкций, который пользователь передает нейросети для выполнения определенной задачи. Он помогает системе лучше понять намерения человека. Например, если нейронная сеть не может ответить на запрос от пользователя, она может так и сгенерировать ответ, что на данный момент не обладает такой информацией. Можно это включить в сам запрос, например, «если ты не знаешь точного ответа или у

тебя недостаточно информации для ответа, то так и напиши, что не могу дать точный ответ».

На самом деле, нейросети обучены таким образом, что они стараются выглядеть полезными для пользователя. Именно вовлечение человека в диалог и его поддержание приводят к тому, что нейронная сеть выдает те ответы, которые устроили бы пользователя.

Можно построить диалог с нейросетью таким образом, чтобы модели анализировали, насколько они уверены в своих ответах, и только после этого их выдавать. В данном случае можно использовать следующий промпт: «Оцени свою уверенность в ответе и объясни почему». Кроме того, важно, чтобы она не генерировала ответ на основе недостаточной информации. Другими словами, обобщая все вышесказанное, нейросети не должны просто догадываться и придумывать информацию, когда не знают ответ. Вместо этого они должны ограничиваться лишь достоверными фактами, то есть для этого тоже можно в запросе указать следующее: «Ответь на вопрос, но каждое утверждение должно быть подкреплено ссылкой на источник или логическим выводом из ранее предоставленной информации».

Современные большие языковые модели (LLM) имеют механизмы оценки достоверности, которые часто подавляются желанием дать полный ответ. Промпт «пробуждает» эти механизмы и использует, чтобы ответы были более честными и надежными ([1]).

Вместо «дай ответ на вопрос» задача будет сформулирована, как «отдели достоверное от недостоверного», что меняет целевую функцию модели в процессе генерации ([1]).

Промпт будет снимать внутреннее давление «всегда знать ответ», позволяя ИИ признавать ошибки без потери «лица» и авторитета. Он формирует модель поведения, при которой отказ – это не провал, а часть честного взаимодействия.

Традиционная модель общения с ИИ и нейросетями неявно поощряет ложное ощущение всезнания. Пользователь спрашивает — нейросеть отвечает, причем почти всегда уверенно и развернуто. Однако это опасно, так как она создает иллюзию разговора с экспертом, когда на самом деле происходит взаимодействие со статистической моделью.

«Протокол достоверности» меняет эту динамику, делая пользователя активным участником процесса верификации, а не потребителем информации. Он устанавливает новый социальный контракт: модель честно признает свои ограничения, а пользователь принимает эти ограничения как неотъемлемую часть технологии, а не как сбой ([1]).

В завершении рассмотрим еще один пример. Предложим нейросети решить линейное уравнение первого порядка: «Реши дифференциальное уравнение $(x + y^2)dy = ydx$ и определи его тип». Нейросеть сначала делит это уравнение на dy , затем на y . Утверждает, что «это линейное дифференциальное уравнение первого порядка относительно x как функции

от y . Оно имеет вид $\frac{dx}{dy} + P(y)x = Q(y)$, где $P(y) = -\frac{1}{y}$ и $Q(y) = y$ ». После

этого используется интегрирующий множитель, умножив на который обе

части уравнения, получается: $\frac{1}{y} \frac{dx}{dy} - \frac{1}{y^2} x = \frac{1}{y}$. «Левая часть является

производной произведения $\frac{d}{dy} \left(\frac{x}{y} \right)$ ». Проинтегрировав обе части по y и

выразив x , получается, что $x = y^2 + Cy$. Однако можно заметить, что в самом начале решения было деление на y , при этом, проверки потеряно это решение или нет, не было. Предполагая, что это нам неизвестно, сделаем запрос: «Оцени свою уверенность в ответе и объясни». Нейросеть утверждает, что полностью уверена в правильности данного ответа, и по пунктам приводит свои объяснения, однако потерянное решение не указывает. Для того чтобы получить окончательный верный ответ, можно

напрямую задать вопрос: «Не потеряны ли решения?». Нейронная сеть соглашается, что действительно $y = 0$ – это тоже решение. Данный пример показывает, что промпт-интерфейс тоже не полностью решает проблему галлюцинаций нейросетей.

Одним из ключевых этапов создания безопасной нейросети является этап «воспитания», где главной проблемой является сверхоптимизация ([2]). В процесс дообучения модель стремится сильнее отличаться от первоначальной версии, чтобы быть максимально полезной. Однако если она отходит слишком далеко, то начинает генерировать бессмысленные, неверные ответы. Главной причиной этой проблемы является то, что первоначальная версия остается неподвижной. «Это все равно, что пытаться научить корабль навигации, заставляя его постоянно оглядываться на порт отправления. Чем дальше он уходит в море, тем менее релевантной становится эта исходная точка». Современные исследования предлагают поступить иначе, в процессе обучения периодически менять саму референтную модель, заменяя ее текущей, улучшенной версией ([2]). Начальную модель следует рассматривать как динамический ориентир, который движется с моделью по мере совершенствования ([2]). Такой подход позволяет уходить от старта дальше, не теряя качество. Модель можно в дальнейшем улучшать, не рискуя сломать ее базовые способности.

Таким образом, использование промпт-интерфейса позволяет решать некоторые проблемы, связанные с галлюцинациями нейросетей: отказ от генерации ответа для нейросети – это не провал, это частная реакция на незнание запроса от пользователя в данный момент. Внедрение промпт-интерфейса улучшит поведение нейросети в сторону правдивости ответов, однако галлюцинации нейросетей остаются одной из ключевых задач в области ИИ, требующих дальнейших исследований для обеспечения полного доверия к ИИ и нейронным сетям. Однако более качественное

«воспитание» языковых моделей напрямую ведет к созданию более умных и безопасных нейросетей. Например, борьба с сверхоптимизацией через внедрение динамического референтного ориентира, когда исходная модель в процессе обучения периодически обновляется на текущую версию, способна сделать нейросеть более безопасной. Это означает меньше ошибок в ответах, более точное изложение, более надежное поведение ИИ в реальных задачах. Хотя даже при решении проблемы галлюцинаций всё-таки, наверное, не стоит полностью доверять нейросетям и ИИ.

Список литературы

1. Галлюцинации моделей текстовых ИИ, и как с ними бороться. URL: <https://habr.com/ru/companies/timeweb/articles/910056/> (дата обращения: 17.11.2025).
2. Исследователи предложили подход к «воспитанию» языковых моделей, уменьшающий количество неуместных или «галлюцинаторных» ответов. URL: <https://naked-science.ru/article/column/issledovateli-predlozhili> (дата обращения: 17.11.2025).
3. Ксенофонов В. В. Нейронные сети // Проблемы науки. 2020. №11 (59). URL: <https://cyberleninka.ru/article/n/neyronnye-seti-1> (дата обращения: 17.11.2025).
4. Кузнецова И. О., Малютов Д. А. Принцип работы и архитектура нейронных сетей // Евразийская интеграция: современные тренды и перспективные направления. 2024. №4. С. 106-111. URL: <https://cyberleninka.ru/article/n/printsip-raboty-i-arhitektura-neyronnyh-setey> (дата обращения: 17.11.2025).
5. Ложь искусственного интеллекта. URL: <https://habr.com/ru/articles/891302/> (дата обращения: 17.11.2025).
6. Фаустова К. И. Нейронные сети: применение сегодня и перспективы развития // Территория науки. 2017. №4. URL: <https://cyberleninka.ru/article/n/neyronnye-seti-primenenie-segodnya-i-perspektivy-razvitiya> (дата обращения: 17.11.2025).

Электронное издание

ЛОБАЧЕВСКИЙ И ХХІ ВЕК

**Материалы XII научно-образовательной студенческой
конференции, посвященной Дню рождения Н.И. Лобачевского**

Компьютерная верстка

А.А. Нурмухаметовой

Гарнитура «Times New Roman, Calibri».