ИЗДАТЕЛЬСТВО грамота

Филологические науки. Вопросы теории и практики
Philology. Theory & Practice

ISSN 2782-4543 (online)
ISSN 1997-2911 (print)

2026. Том 19. Выпуск 9 | 2026. Volume 19. Issue 9
Материалы журнала доступны на сайте (articles and issues available at): philology-journal.ru

RU

Сравнительное экспериментальное исследование интерпретации окказиональных фразеологических единиц нейросетевыми моделями на материале русско- и англоязычного медиадискурсов

Хакимуллина Д. Ф., Демираг Д. Н.

Аннотация. Представленная статья посвящена проблеме галлюцинаций больших языковых моделей (LLM) при работе с комплексными языковыми единицами – окказиональными фразеологизмами. Несмотря на большую распространенность нейросетей в лингвистическом поле, их ответы нередко демонстрируют семантическую несостоятельность, при этом оставаясь убедительными «внешне». Этот феномен именуется галлюцинацией искусственного интеллекта (ИИ). Целью исследования является разработка типологии галлюцинаций ИИ при интерпретации окказиональных фразеологических единиц в русско- и англоязычном медиадискурсах. Научная новизна исследования состоит в том, что в работе впервые на русском и английском материале разработана и эмпирически обоснована классификация галлюцинаций, специфичная для окказиональных фразеологических единиц. Полученные результаты показали, что русский язык представляет для LLM значительно большую сложность, чем английский (средняя точность 56% против 68%), а типология галлюцинаций имеет ярко выраженную кросс-языковую специфику: на английском материале доминирует буквализация (40%), на русском – ошибка прототипизации (35%) и культурная псевдокомпетенция (25%). Предложенная типология может быть использована при разработке тестовых наборов для оценки LLM и в практике преподавания лингвистических дисциплин.

EN

Comparative experimental study of the interpretation of occasional phraseological units by neural network models based on Russian- and English-language media discourses

D. F. Khakimullina, D. N. Demirag

Abstract. The presented article is devoted to the problem of hallucinations in large language models (LLMs) when processing complex linguistic units – occasional phraseological units. Despite the widespread prevalence of neural networks in the linguistic field, their responses often demonstrate semantic inadequacy while remaining convincing "on the surface." This phenomenon is referred to as AI hallucination. The aim of the research is to develop a typology of AI hallucinations in the interpretation of occasional phraseological units in Russian- and English-language media discourses. The scientific novelty of the study lies in the fact that, for the first time, a classification of hallucinations specific to occasional phraseological units has been developed and empirically substantiated using Russian and English material. The results obtained showed that the Russian language presents significantly greater difficulty for LLMs than English (an average accuracy of 56% versus 68%), and the typology of hallucinations exhibits a pronounced cross-linguistic specificity: literalization dominates in the English material (40%), whereas prototypicality error (35%) and cultural pseudo-competence (25%) prevail in the Russian material. The proposed typology can be used in developing test suites for evaluating LLMs and in the practice of teaching linguistic disciplines.

Введение

Большие языковые модели (Large language models – LLM) регулярно используются для задач различного характера. Их повсеместное внедрение в повседневные бытовые, лингвистические, переводческие и редакторские

Научная статья (original research article) | https://doi.org/10.30853/phil20260435

практики показывает, что, несмотря на очевидные удобство, скорость и, зачастую, высокое качество ответов, которые генерируют нейросети, нередко их ответы могут не иметь ничего общего с «правильным», а иногда и «разумным» (Шалак, 2024; Wang, Singh, Michael et al., 2018). Несмотря на то, что результаты запроса обычно выглядят логично и «осмысленно», за ними могут маскироваться бессмысленные ответы. Более того, нейросеть, генерируя такие ответы, не предупреждает пользователя о возможной недостоверности, тем самым не мотивирует его на критическое отношение к результату. В целом, модели часто дают грамматически верные, аргументированные и убедительные интерпретации даже сложных языковых единиц, однако экспертная проверка нередко признает такие результаты несостоятельными и немотивированными (Кружилина, 2023). Подобные генерации именуются галлюцинацией. Данное явление, когда форма ответов имитирует осознанность, но содержание этого не отражает, имеет комплексную структуру (Sun, Sheng, Zhou et al., 2024). Именно этому феномену посвящено представленное исследование.

Галлюцинация нейросетей имеет настолько высокую частотность, что в 2023 году этот термин, применимый к ИИ, стал словом года по версии Cambridge English Dictionary. Проблема различения «реального» понимания и имитации понимания стала центральной в прикладных лингвистических исследованиях (Krakauer, Krakauer, Mitchell, 2025; Bottazzi, Ferrario, 2025, Bender, Koller, 2020). Для фиксирования особенностей галлюцинаций наиболее подходящими кажутся окказиональные фразеологические единицы (ОФЕ). Диагностический потенциал ОФЕ обусловлен следующими свойствами: во-первых, ОФЕ не фиксируются в обучающих массивах текстов; во-вторых, для работы с такими комплексными единицами требуется инференция с учетом контекста и восстановлением узуальной формы. Активное внедрение ИИ в лингвистические, переводческие, редакторские практики постепенно приводит к большому уровню доверия пользователя к постоянно совершенствующимся моделям несмотря на частотность галлюцинации. Актуальность исследования обусловлена необходимостью системного анализа способности больших языковых моделей анализировать сложные языковые единицы для определения границ их применения в лингвистическом поле, поскольку без подобных данных невозможно оценить, насколько корректно нейросетевые модели работают с семантически сложными или культурно-обусловленными единицами языка.

Для достижения цели исследования был сформулирован ряд задач:

1. Разработка классификации галлюцинаций при интерпретации ОФЕ.
2. Сравнение ответов трех нейросетей по ряду предлагаемых в работе параметров.
3. Сопоставление полученных данных с точки зрения кросс-языковых различий.

Теоретическую базу исследования составили следующие работы.

В области фразеологии ключевыми для представленного исследования стали труды Е. Ю. Семушиной и Э. Э. Валеевой (2025), рассматривающих способы образований окказиональной расширенной фразеологической метафоры, что составило основу для выявления типов трансформации ОФЕ в работе; И. Ю. Третьяковой (2011) в вопросах окказиональной фразеологии. В области ИИ в лингвистике – H. Liu, W. Xue, Y. Chen (Liu, Xue, Chen et al., 2024), Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung (2023), V. Rawte, S. Chakraborty, A. Pathak, A. Sarkar, S. M. T. I. Tonmoy, A. Chadha, A. Sheth, A. Das (2023). Z. Ji, N. Lee, R. Frieske и соавторы предлагают уникальные классификации типов галлюцинаций ИИ, которые легли в основу представленной в работе типологии.

Материалом для исследования послужил авторский корпус из 150 газетных заголовков с содержанием ОФЕ (75 на русском языке и 75 на английском языке). Материал был собран из газет «Коммерсантъ», «Известия», «Аргументы и Факты», “The Sun”, “The Guardian”, “The Times”. Дата публикации статей: не ранее 2020 года для сохранения языковой актуальности. В эксперименте участвовали 3 нейросети: DeepSeek (Китай), Perplexity (США), YandexGPT (Россия). Выбор обуславливается доступностью всех трех нейросетей на территории РФ и возможностью их использования на бесплатной основе. Данные критерии делают эти нейросети доступными для широкой аудитории, а потому вызывают исследовательский интерес. Разные страны производства моделей предполагают разницу в обучающих корпусах. Выбор моделей Perplexity и YandexGPT продиктован привязкой к исследуемым языкам, в то время как LLM DeepSeek представлена как универсальная модель.

Для проведения исследования были использованы следующие методы: метод сплошной выборки при сборе авторского корпуса газетных заголовков, метод систематизации языкового материала, элементы лингвокогнитивного моделирования при описании типологии галлюцинаций ИИ, сравнительно-сопоставительный метод при выделении универсальных и уникальных для рассматриваемых языков особенностей галлюцинации.

Теоретическая значимость представленной работы заключается в получении уникальных эмпирических данных в рамках компаративного исследования. Работа вносит определенный вклад в понимание того, как морфологическая и семантическая сложность языка влияет на работу LLM.

Практическая значимость исследования заключается в возможности использования полученных эмпирических результатов и теоретических уточнений типологии галлюцинаций ИИ при работе с комплексными языковыми единицами в разработке теоретических и практических курсов, посвященных вопросам лингвистики и ИИ, теории и практики перевода. Представленные данные могут быть применены при разработке тестовых наборов для оценки работы больших языковых моделей.

Обсуждения и результаты

Учитывая, что понятие галлюцинации ИИ довольно новое и современное, научное сообщество еще не определило единый подход к этому термину. Часто галлюцинация ИИ имеет схожее понимание с термином

галлюцинации в целом, который в первую очередь используется в психологии, из-за схожести обоих явлений. Галлюцинация ИИ тесно связана с философией ИИ, а точнее, с понятием истинности и обоснованности. Особенность галлюцинации с этой точки зрения состоит в том, что модель не «лжет». LLM зачастую ошибается «искренне», с полной уверенностью в своей правоте. У ИИ нет критического мышления, а потому нет и умысла во введении «собеседника» в заблуждение. Потому галлюцинация ИИ является феноменом не этическим, а эпистемическим (Шевченко, 2025).

Ученые считают, что галлюцинация ИИ – это «случаи, в которых ИИ генерирует вымышленные, ошибочные или необоснованные ответы» (Ji, Lee, Frieske et al., 2023, p. 4). Основная проблема галлюцинации ИИ как явления заключается в том, что она не связана с недостаточной обученностью машины (Rawte, Chakraborty, Pathak et al., 2023; Ji, Lee, Frieske et al., 2023) Это – структурный дефект. Согласно В. Барасси, нейросеть обучается на массивах текстов, «в которых уже заложены искажения, иерхахии и умолчания» (Barassi, 2024, p. 7). Потому галлюцинация – это не просто «фантазия» ИИ, а отражение этих скрытых структур. Данное утверждение не лишено логики и смысла, однако вопрос, почему некоторые ответы ИИ выглядят абсолютно бессмысленно и по сути являются «ответом ради ответа», данная позиция не решает. И все же мысль о том, что галлюцинация является не недоработкой, а особенностью устройства когнитивных операций, остается актуальной.

Принципиальное отличие галлюцинации от обычной ошибки (что является причиной повышенного интереса научного сообщества к этому явлению) заключается в том, что она часто «маскируется» под правильный ответ. Без принципиальной критичности к результату пользователь может принять галлюцинацию за истину.

Разработка проблемы галлюцинации ИИ не обходится без классификации данного явления. Этим занимаются ученые различных направлений исследований (Айсин, Шамардина, 2025; Rawte, Sheth, Das, 2023; Liu, Xue, Chen et al., 2024). Наиболее лаконичной и легко имплементируемой в лингвистическое поле кажется классификация Z. Ji, N. Lee, R. Frieske (Ji, Lee, Frieske et al., 2023):

1. Внутренняя галлюцинация (intrinsic hallucination). Такая галлюцинация противоречит контексту или встроенным в модель фактам. Модель ошибается.

2. Внешняя галлюцинация (extrinsic hallucination). Данный тип галлюцинации не может быть ни подтвержден, ни опровергнут исходным контекстом. Модель «додумывает».

Вслед за указанными выше авторами классификации галлюцинаций ИИ мы, учитывая лингвистическую специфику исследования, видим необходимым уточнить ее. Авторам исследования кажется логичным выдвинуть более узкую типологию галлюцинаций ИИ при работе с комплексными языковыми единицами (такими как ОФЕ):

1. Внутренняя галлюцинация

1.1. Буквализация. Данный тип классификации предполагает игнорирование метафоричности и образности языковой единицы (в том числе и ОФЕ). Значение воспринимается буквально, через прямой смысл компонентов.

1.2. Ошибка прототипизации. Модель понимает, что перед ней ОФЕ, но, даже при верной интерпретации, либо восстанавливает узуальную форму неверно (галлюцинирует и создает новую ФЕ), либо связывает ОФЕ с неверной узуальной формой.

2. Внешняя галлюцинация

2.1. Культурная псевдокомпетенция. Модель генерирует уверенное объяснение, добавляя вымышленные или ошибочные отсылки к культурным, историческим или иным деталям.

2.2. Инференциальная ошибка. Модель представляет «логичный» связный ответ, но не улавливает идиоматичность.

Как было сказано ранее, для проверки работы ИИ со сложными языковыми единицами авторами было принято решение использовать ОФЕ. Данный выбор мотивирован тем, что ОФЕ почти наверняка отсутствуют в обучающих корпусах LLM из-за своей уникальности. При встрече с ОФЕ машина не может опираться исключительно на статистические данные, ей приходится строить интерпретацию на основе контекста или внутренних механизмов. Так ОФЕ как диагностический материал позволяет увидеть, насколько модель «понимает» язык и может работать с лингвокреативностью, а не просто повторять статистически выверенные паттерны.

ОФЕ – это комплексная языковая единица. Несмотря на окказиональные преобразования узуальной формы она сохраняет материальные и семантические связи с исходной ФЕ (Петрова, 1984; Батурова, 2013). Высокая устойчивость и воспроизводимость ФЕ делает ОФЕ ценным языковым материалом, так как «вмешательство» в узуальную форму выглядит не как ошибка, а как реализация лингвокреативного потенциала. Высокая образность ОФЕ и наложение узуальной и окказиональной семантик друг на друга позволяют признать данную единицу уникальным языковым материалом (Семушина, Валеева, 2025). Легкое соотнесение окказиональной и узуальной форм в сознании адресата сообщения позволяет автору текста реализовывать игру слов, наслаивать смыслы для достижения желаемого эффекта. И. Ю Третьякова называет окказиональными такие ФЕ, которые «реализуют авторскую творческую мысль» в определенном контекстном окружении (2011, с. 161). Для работы с ИИ в рамках экспериментального исследования авторами были сформулированы 2 промпта для проверки качества работы нейросети с ОФЕ на двух уровнях: 1) интерпретация ОФЕ с опорой на контекст; 2) восстановление прототипа (промпты были написаны на языке контекста для исключения ошибок, связанных с переключением языка). Отметим, что в вопросе на реконструкцию узуальной формы нейросети предлагался ответ «Я не знаю» для исключения генерации ответа-галлюцинации. Работа с ИИ на разных уровнях позволила глубже рассмотреть вопрос особенностей восприятия сложных языковых единиц нейросетями и идентифицировать особенности галлюцинации LLM, которые могли бы остаться незамеченными при анализе способности интерпретации или прототипизации изолированно.

Промпты были представлены нейросетям с сохранением изолированных сессий для гарантии отсутствия влияния предыдущих запросов на ответ.

Во время исследования авторы зафиксировали все 900 результатов взаимодействия с нейросетью. Каждый из них был оценен по трехбалльной шкале: 2 – полностью верно; 1 – частично верно; 0 – неверно. Галлюцинацией считался неверный ответ нейросети, «замаскированный» под верный и логически обоснованный. Ответ «Я не знаю» оценивался в 0 баллов, но не рассматривался как галлюцинация. Процент галлюцинаций, зафиксированный на каждом из двух уровней, представлен в Таблице 1.

Таблица 1. *Количественные показатели галлюцинаций ИИ*

Язык	Задание	DeepSeek	Perplexity	YandexGPT
Русский	Интерпретация	14%	22%	17%
	Прототипизация	20%	22%	33%
Английский	Интерпретация	12%	15%	17%
	Прототипизация	12%	19%	15%

Отметим, что наиболее высокий уровень галлюцинаций наблюдается на материале русскоязычных ОФЕ в задании на прототипизацию, а самый низкий – на англоязычном материале в задании на интерпретацию. В целом интерпретация имеет более высокий процент успешности благодаря возможности опоры на контекст. Отсутствие «понимания» ОФЕ может быть компенсировано окружением.

Наибольший интерес представляет классификация «ошибок» ИИ с опорой на предложенную выше типологию.

Буквализация. Нейросеть нередко действительно бывает «слепа» к метафоричности и образности ОФЕ. Так, контекст «ОСТАНУТСЯ БЕЗ ШТАНОВ? “Глория Джинс” может закрыть самый большой магазин» (https://rostov.aif.ru/society/ostanutsya-bez-shtanov-gloriya-dzhins-mozhet-zakryt-samyy-bolshoy-magazin) был интерпретирован LLM DeepSeek следующим образом: «*Фраза “ОСТАНУТСЯ БЕЗ ШТАНОВ” в данном контексте означает, что покупатели или сотрудники магазина “Глория Джинс” могут лишиться возможности приобрести или получить джинсы (штаны), если самый большой магазин компании закроется*» (DeepSeek). Привязка метафорческой семантики лишения к штанам излишне буквализирует ситуацию. Семантическая игра в целом игнорируется.

Ошибка прототипизации. Восстановление узуальной формы действительно требует от модели серьезной работы над тем языковым материалом, на котором она обучалась. Эта работа зачастую выходит за рамки статистического соотнесения ЛЕ. Одним из наиболее ярких примеров данного типа галлюцинаций является контекст «КОД ЧУМЫ И КОД ТЮРЬМЫ: прокуратура взялась за наркорекламщиков» (https://iz.ru/1529099/ivan-petrov/kod-chumy-i-kod-tiurmy-prokuratura-vzialas-za-narkoreklamshchikov). Определение прототипа данной ОФЕ стало сложностью для двух моделей из трех. Так, модель DeepSeek представила следующий прототип: «*“Кодекс чести” (изменено на “код чумы”) и “тюремный кодекс” (изменено на “код тюрьмы”); способ изменения – замена частей исходных устойчивых сочетаний (“кодекс” → “код”, “чести” → “чумы”, а также контаминация/параллелизм с “код тюрьмы” вместо “тюремный кодекс/закон”)*» (DeepSeek). Модель Perplexity ответила следующим образом «*Исходная идиома – “Кодекс чести и кодекс тюрьмы” (или более известная вариация “Кодекс чести и осуждённых”), но здесь она изменена заменой слова “кодекс” на “код” и слова “чести” на “чумы”, что создаёт игру слов*» (Perplexity). Очевидно, что узуальная ФЕ «от сумы и (да) от тюрьмы не открестишься/зарекаться» была полностью проигнорирована обеими моделями. Предположим, что это связано с меньшим количеством русскоязычных текстов, участвующих в их обучении, так как нейросеть российского производства YandexGPT не имела проблем с указанной ОФЕ.

Культурная псевдокомпетенция. Отсылки к элементам культурного кода нации не всегда успешно считываются нейросетями. Так, контекст с ОФЕ «ТАМ СМУГЛЯНКА-МОЛДОВАНКА СОБРАЛА ЭЛЕКТОРАТ. Что будет после того, как Майя Санду выиграла выборы…» (https://www.kommersant.ru/doc/4574831) был прототипизирован моделью Perplexity следующим образом: «*Исходная идиома – “Смуглянка-молдаванка собрала вишню” (из песни)*» (Perplexity). В целом, примеры культурной слепоты оказались наиболее характерны для контекстов на русском языке. Вероятно, это связано с большим влиянием искусства на русский язык. Тесное переплетение языка и кинематографа, литературы, музыки приводит к тому, что культурная обусловленность фразеологического фонда повышается.

Инференциальная ошибка. Непонимание идиоматичности является нередкой ошибкой ИИ при работе с ОФЕ. Контекст “TALKING HORSES: Sunday night races may result in serious jockey burnout” (https://www.theguardian.com/sport/blog/2024/jan/08/talking-horses-sunday-night-races-may-result-in-serious-jockey-burnout-horse-racing-tips) / «Говорящие лошади (буквальное) / Разговор о лошадях (буквальное) / Лошадиное чутье (метафорическое): воскресные ночные скачки могут привести к серьезному выгоранию жокеев» (здесь и далее перевод выполнен автором статьи. – *Д. Х.*) был интерпретирован следующим образом: *“The phrase "TALKING HORSES" suggests that the horses themselves are communicating... if horses could talk, they might have something to say about it.” / «Фраза предполагает, что лошади общаются… если бы лошади умели говорить, они бы могли что-то сказать по этому поводу*» (DeepSeek). Буквальное понимание ОФЕ и непризнание ее образности очевидно. Невосприимчивость к идиоматичности прослеживается и при интерпретации ОФЕ «Погода в DOOM’е» (https://iz.ru/1889879/anton-belyi/pogoda-v-doome/novuiu-chast-kultovoi-serii-boevikov-doom-pereveli-na-russkii-iazyk). LLM DeepSeek представляет следующую интерпретацию данной ОФЕ: «*Фраза иронично обыгрывает известный вопрос “какая погода в Лондоне?” …*» (DeepSeek). Отсутствие связи с известной строчкой из песни в результате приводит к уверенному, аргументированному, но неверному ответу.

Ознакомиться с количественными показателями распределения галлюцинаций ИИ при работе с ОФЕ в зависимости от языка можно в Таблице 2.

Таблица 2. *Распределение галлюцинаций нейросетей по типам*

Тип галлюцинации	Русский язык	Английский язык
Буквализация	20%	40%
Ошибка прототипизации	35%	30%
Культурная псевдокомпетенция	25%	5%
Инференциальная ошибка	20%	25%

Анализ показал, что все три LLM демонстрируют более низкую точность при работе с русским языком. Наибольший разрыв зафиксирован в задании на прототипизацию, что можно объяснить большей культурной спецификой фразеологического фонда русского языка.

В вопросе галлюцинаций можно отметить, что на английском материале доминирует буквализация значения. Предполагаем, что это связано с высоким уровнем привязки идиом (в том числе и окказионально трансформированных) к реальным образам. В русском языке лидируют такие типы галлюцинаций, как ошибка прототипизации и культурная псевдокомпетенция. При работе с русскими контекстами модели либо не могут восстановить прототип, либо демонстрируют немотивированную уверенность в ошибочных культурных отсылках.

Заключение

Проведенное исследование позволило авторам прийти к следующим выводам.

В результате исследования авторами была разработана и обоснована типология галлюцинаций, специфичная для работы ИИ с комплексными языковыми единицами. Предложенная классификация базируется на уже предложенной двухуровневой классификации галлюцинаций ИИ (внутренние и внешние) и дополнена авторами для отражения специфики работы больших языковых моделей с уникальными языковыми единицами. Она состоит из двух видов внутренней (буквализация и ошибка прототипизации) и двух видов внешней (культурная псевдокомпетенция, инференциальная ошибка) галлюцинаций ИИ. Дополненная классификация может быть использована в дальнейших исследованиях для выявления типов ошибок ИИ при работе с обработкой естественного живого языка и при разработке тестового материала для коррекции работы больших языковых моделей.

Сравнительный анализ ответов трех нейросетей проводился по ряду параметров: точность ответов (была применена трехбалльная система оценки); распределение неверных ответов по типу галлюцинации; распределение по уровню анализа (уровень интерпретации и уровень прототипизации).

Нейросеть DeepSeek продемонстрировала наилучший результат по всем указанным параметрам. Ее общая точность составила 66%, а высокая успешность в задании на прототипизацию подтверждает отсутствие опоры этой модели исключительно на статистические данные, выведенные при анализе массивов обучающих текстов. Однако данная модель уязвима к инференциальным ошибкам и буквализации на английском материале. Второе место по количеству галлюцинаций занимает LLM YandexGPT. Российская модель продемонстрировала более высокий процент успешности анализа англоязычных контекстов, нежели русскоязычных. Эта модель уступает DeepSeek в вопросе прототипизации, данный тип галлюцинаций частотен. Помимо этого, модель имеет проблемы с культурной обусловленностью. Модель Perplexity является лидером по количеству галлюцинаций при анализе авторского корпуса. Данная модель продемонстрировала самый высокий процент полностью неверных ответов (оцененных в 0 баллов), особенно в заданиях на прототипизацию русских ОФЕ. Эта языковая модель имеет трудности с культурной псевдокомпетенцией. Часто ответы абсурдны и алогичны.

Сопоставление результатов на русском и английском языках выявило устойчивые кросс-языковые различия.

Точность ответов: русский язык оказался сложнее для всех трех моделей. Наибольшую проблему представило задание на восстановление узуальной формы ФЕ. Этот факт объясняется большей культурной обусловленностью и высокой идиоматичностью фразеологического фонда русского языка. Вероятно, ФЕ менее распространены в обучающих массивах текстов на русском языке.

Тип галлюцинации ИИ: буквализация как тип галлюцинации ИИ лидирует при работе моделей с английскими ОФЕ (около 40% всех ошибок), на русском же языке ошибка прототипизации и культурная псевдокомпетенция (35% и 25% соответственно) имеют больший процент репрезентации.

Культурная специфика: на основе проанализированных данных мы можем утверждать, что культурная псевдокомпетенция – это уникальный для русского языка тип галлюцинации. Прецедентные тексты, фольклор, советская и постсоветская культура оказали большое влияние на фразеологический фонд русского языка, а потому как узуальные, так и окказиональные ФЕ могут представлять для нейросетей, базирующих свою работу скорее на статистике, чем на «осмыслении», определенную зону риска.

Таким образом, представленное исследование подтвердило, что галлюцинации LLM при работе с такими комплексными языковыми единицами представляют собой системную проблему, обусловленную частичной непригодностью архитектурных особенностей нейросетей для анализа лингвокреативности и окказиональных единиц в целом, так как они не поддаются статистическому анализу. Предложенная в работе типология галлюцинаций ИИ позволяет дифференцировать и упорядочить ошибки ИИ и диагностировать трудности

нейросети. Контрастивный языковой анализ показал, что русский язык представляет большую сложность для языковых моделей, что требует пересмотра количества и качества массива обучающих текстов на русском языке. Результаты анализа в целом демонстрируют острую необходимость разработки методов снижения галлюцинаций для сложных языковых единиц.

Перспективы дальнейшего исследования определяются спецификой работы с большими языковыми моделями. Возможно как расширение исследуемого корпуса (до минимальных 300 единиц) для большей детализации типологии, так и увеличение количества моделей, тестируемых в рамках эксперимента, для наиболее объективного анализа особенностей интерпретации сложных языковых единиц нейросетями. Особый интерес может представлять включение в работу контрольной группы носителей исследуемых языков для сравнения их интерпретации ОФЕ с большими языковыми моделями. Кроме того, перспективным направлением может стать применение разработанной классификации в отношении анализа других комплексных языковых единиц (паремии, метафоры).

Материалы исследования | Research materials

1. Аргументы и Факты. https://aif.ru/
2. Известия. https://iz.ru/
3. Коммерсантъ. https://www.kommersant.ru/
4. Cambridge English Dictionary. https://dictionary.cambridge.org/
5. The Guardian. https://www.theguardian.com/international
6. The Sun. https://www.thesun.co.uk/
7. The Times. https://www.thetimes.com/us

Источники | References

1. Айсин Т. Р., Шамардина Т. В. Детекция галлюцинаций на основе внутренних состояний больших языковых моделей // Электронные библиотеки. 2025. Т. 28. Вып. 6. https://doi.org/10.26907/1562-5419-2025-28-6-1282-1305
2. Батурова Н. И. Окказиональный фразеологизм как языковое явление // Известия Тульского государственного университета. Гуманитарные науки. 2013. № 2.
3. Кружилина Т. В. К вопросу о моделировании процесса понимания текста с помощью систем искусственного интеллекта // Известия Юго-Западного государственного университета. 2023. Т. 27. № 4. https://doi.org/10.21869/2223-1560-2023-27-4-98-116
4. Петрова С. И. Окказиональные фразеологизмы в немецкой художественной речи (структурно-семантическая и смысловая характеристики): дисс. ... к. филол. н. М., 1984.
5. Семушина Е. Ю., Валеева Э. Э. Особенности образования окказиональной расширенной фразеологической метафоры (на материале английского и русского языков) // Филологические науки. Вопросы теории и практики. 2025. № 3. https://doi.org/10.30853/phil20250123
6. Третьякова И. Ю. Окказиональные фразеологизмы и контекст // Ярославский педагогический вестник. 2011. № 2.
7. Шалак В. И. Избавление от иллюзий ИИ на примере ChatGPT // Technology and Language. 2024. № 5 (2). https://doi.org/10.48417/technolang.2024.02.03
8. Шевченко А. А. «Галлюцинации» ИИ как новая форма эпистемической ошибки // Respublica Literaria. 2025. Т. 6. № 4. https://doi.org/10.47850/RL.2025.6.4.93-98
9. Barassi V. Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI // Harvard Data Science Review. 2024. Special Issue 5.
10. Bender E. M., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020.
11. Bottazzi G. E., Ferrario R. The Bewitching AI: The Illusion of Communication with Large Language Models // Philosophy & Technology. 2025. Vol. 38. Iss. 2 (61). https://doi.org/10.1007/s13347-025-00893-6
12. Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y. J., Madotto A., Fung P. Survey of Hallucination in Natural Language Generation // ACM Computing Surveys. 2023. Vol. 55. No. 12. https://doi.org/10.1145/3571730
13. Krakauer D. C., Krakauer J. W., Mitchell M. Large Language Models and Emergence: A Complex Systems Perspective // arXiv preprint. 2025. https://doi.org/10.48550/arXiv.2506.11135
14. Liu H., Xue W., Chen Y., Chen D., Zhao X., Wang K., Hou L., Li R., Peng W. A Survey on Hallucination in Large Vision-Language Models // arXiv preprint. 2024. https://doi.org/10.20944/preprints202510.0540.v1
15. Rawte V., Chakraborty S., Pathak A., Sarkar A., Tonmoy S. M. T. I., Chadha A., Sheth A., Das A. The Troubling Emergence of Hallucination in Large Language Models – An Extensive Definition, Quantification, and Prescriptive Remediations // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023. https://doi.org/10.18653/v1/2023.emnlp-main.155
16. Rawte V., Sheth A., Das A. A Survey of Hallucination in Large Foundation Models // arXiv preprint. 2023. https://doi.org/10.48550/arXiv.2309.05922

17. Sun Y., Sheng D., Zhou Z., Yifei W. AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content // Humanities and Social Sciences Communications. 2024. Vol. 11. Article 1278. https://doi.org/10.1057/s41599-024-03811-x

18. Wang A., Singh A., Michael J., Hill F., Levy O., Bowman S. GLUE: A multi-task benchmark and analysis platform for natural language understanding // Proc. of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2018. https://doi.org/10.18653/v1/W18-5446

Информация об авторах | Author information

RU

Хакимуллина Диана Фаридовна[1]
Демираг Диана Няилевна[2], д. филол. н., проф.
[1, 2] Казанский федеральный университет

EN

Diana Faridovna Khakimullina[1]
Diana Nyailevna Demirag[2], Dr
[1, 2] Kazan Federal University

[1] *attenxatusha@gmail.com,* [2] *dianadi@bk.ru*

Информация о статье | About this article

Дата поступления рукописи (received): 19.08.2026; опубликовано online (published online): 11.09.2026.

Ключевые слова (keywords): галлюцинации ИИ; окказиональные фразеологические единицы; большие языковые модели; медиадискурс; фразеологическая трансформация; AI hallucinations; occasional phraseological units; large language models; media discourse; phraseological transformation.