УДК 316.77

ГАЛЛЮЦИНАЦИИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И ИХ ВЛИЯНИЕ НА НАУКУ, МЕДИА И ОБЩЕСТВЕННОЕ ДОВЕРИЕ К ИИ

Ищук Д.А. [1]
Научный руководитель – аналитик Центра научной коммуникации, ассистент Института международного развития и партнерства Савченко А.В. [1]
[1]Университет ИТМО
doraichschuk@yandex.ru

Введение

В 2025 году системы искусственного интеллекта (ИИ), прежде всего генеративные модели, заняли заметное место в публичном коммуникативном пространстве и широко используются в научных исследованиях, образовании, журналистике, медиаиндустрии и повседневных практиках пользователей. На этом фоне усиливаются дискуссии о качестве и надежности ИИ, а также об особенностях взаимодействия человека с нейросетями, что связано не только с широким распространением этих технологий, но и с их влиянием на науку, медиа и повседневные практики.

Особое внимание уделяется феномену галлюцинаций больших языковых моделей – генерированию правдоподобных, но фактически ложных утверждений. Галлюцинации рассматриваются не только как технический сбой, но и как новая форма эпистемической неточности, отличающаяся от человеческой дезинформации отсутствием намерения обманывать [1].

Публичное обсуждение этой проблемы формируется на разных уровнях дискурса. В научных публикациях фокус делают на причинах и механизмах ошибок; в научно-популярных текстах – на доступном объяснении и практических рекомендациях пользователям; в массовых медиа галлюцинации ИИ нередко обретают сенсационный или ироничный характер. Это приводит к фрагментированному представлению о природе галлюцинаций и влияет на уровень доверия к ИИ [2].

Основная часть

Анализ публикаций 2025 года по теме галлюцинаций больших языковых моделей (LLM) выявил три ключевых направления современных исследований: технические причины возникновения и методы минимизации; последствия для научной практики (воспроизводимость, экспертные решения); влияние на медиапространство и общественное доверие к ИИ.

1) Влияние на науку. Галлюцинации подрывают воспроизводимость исследований и экспертные решения. В научных публикациях отмечается, что LLM генерируют ложные цитаты и искажают данные. Исследования показывают систематические ошибки в обработке научных текстов, что усложняет использование ИИ в экспертных задачах. Ученые фиксируют рост настороженности: 64% исследователей выражают беспокойство по поводу галлюцинаций, хотя активно применяют LLM в работе [3]. С опытом доверия становится меньше – ИИ воспринимается как инструмент, требующий строгой проверки.

2) Влияние на медиа. В журналистике галлюцинации приводят к искажению новостной повестки и необходимости пересмотра стандартов фактчекинга. Международные исследования выявили значительную долю фактических ошибок в ответах ИИ на новостные запросы, что подрывает доверие к автоматизированному контенту. Медиа вынуждены вводить дополнительные проверки и отказываться от безоговорочного использования LLM для генерации текстов [4].

3) Влияние на общественное доверие. Галлюцинации формируют амбивалентное

отношение к ИИ: от тревоги за рабочие места, приватность и контроль до восхищения успехами. Массовые медиа драматизируют случаи ошибок, усиливая страх, а мемы превращают абсурдные галлюцинации в иронию. Глобальные опросы показывают рост осторожности: люди признают ограничения технологий, но сохраняют фрагментированные представления. Разработчики отмечают, что методы обучения нейросетей поощряют угадывание вместо честного признания незнания, что усиливает кризис доверия [5].

Выводы

Галлюцинации больших языковых моделей (LLM) в 2025 году получили широкое обсуждение на всех уровнях публичного дискурса как системная проблема, затрагивающая науку, медиа и общество. Все типы обсуждения фиксируют отрезвление: технологии признают полезными, но требующими строгого контроля из-за рисков.

Для снижения рисков необходимы технические меры (улучшение архитектуры моделей, повышение их прозрачности) и просветительские форматы для пользователей. Перспективно сотрудничество науки и медиа для формирования взвешенного образа ИИ.

Литература

1. Шевченко А.А. «Галлюцинации» ИИ как новая форма эпистемической ошибки // Respublica Literaria. – 2025. – Т. 6, № 4. – С. 93–98.
2. New sources of inaccuracy? A conceptual framework for studying AI hallucinations [Электронный ресурс] // Harvard Kennedy School Misinformation Review. – 2025. – URL: https://misinforeview.hks.harvard.edu/article/new-sources-of-inaccuracy-a-conceptual-framework-for-studying-ai-hallucinations/ (дата обращения: 07.01.2026).
3. European Broadcasting Union; BBC. AI’s systemic distortion of news is consistent across languages and territories: international study by public service broadcasters [Электронный ресурс]. – URL: https://www.ebu.ch/news/2025/10/ai-s-systemic-distortion-of-news-is-consistent-across-languages-and-territories-international-study-by-public-service-broadcaste (дата обращения: 07.01.2026).
4. When AI gets it wrong: addressing AI hallucinations and bias [Электронный ресурс] // MIT Sloan. – URL: https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/ (дата обращения: 07.01.2026).
5. OpenAI. Why language models hallucinate [Электронный ресурс]. – URL: https://openai.com/index/why-language-models-hallucinate/ (дата обращения: 07.01.2026).
6. Wiley. AI in research: global study [Электронный ресурс]. – URL: https://www.wiley.com/content/wiley-com/na/us/en/about-us/ai-resources/ai-study.html (дата обращения: 07.01.2026).

Ищук Д. А. _____

Савченко А.В. _____