

РОЛЬ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В АНАЛИЗЕ ЯЗЫКОВОЙ И КУЛЬТУРНОЙ КАРТИНЫ МИРА

Работа посвящена исследованию роли больших языковых моделей в анализе языковой и культурной картины мира. Обосновывается нецелесообразность использования сгенерированных текстов для анализа языковой картины мира, и предлагается альтернатива – исследование требований к моделям как проявлений культурных и этических норм и ценностей современного общества.

Ключевые слова: большие языковые модели, языковая картина мира, культурная картина мира.

N.S. Yakimova, Samara University

THE ROLE OF LARGE LANGUAGE MODELS IN THE ANALYSIS OF THE LANGUAGE AND CULTURAL WORLDVIEW

The work is devoted to the study of the role of large language models in the analysis of the linguistic and cultural picture of the world. The author substantiates the inexpediency of using the generated texts to analyze the linguistic picture of the world and suggests an alternative – the study of requirements for models as manifestations of cultural and ethical norms and values of modern society.

Keywords: large language models (LLM), linguistic worldview, cultural worldview.

Язык – не только средство общения, но и источник обобщенных знаний о мире и человеке. Еще в XVIII веке В. Фон Гумбольдт писал о внутренней форме языка, «языковом мировидении»: язык формирует «внутреннюю форму» мышления, определяя восприятие мира носителями языка [1]. Эта идея получила развитие в гипотезе Сепира–Уорфа, утверждающей, что структура языка влияет на мышление и мировосприятие [2; 3]. В XX веке окончательно оформилась мысль о том, что в текстах содержится «некая картина мира» [4], появился термин *языковая картина мира*, который может пониматься как представления о мире, зафиксированные в текстах, написанных на определенном языке.

Большие языковые модели (БЯМ) генерируют тексты. И можно предположить, что тексты, созданные БЯМ, можно использовать как материал для анализа

языковой картины мира. Однако это может быть сопряжено с рядом методологических проблем:

1. Объем и разнообразие данных. Количество сгенерированных текстов огромно, что затрудняет выборку и анализ.

2. Качество и стабильность моделей. Модели развиваются, постоянно обновляются. Например, OpenAI в апреле 2025 года выпустила GPT-4.1, а также представила рассуждающие модели o3 и o4-mini; о разработке модели YandexGPT было объявлено в феврале 2023 года, а феврале 2025-го компания «Яндекс» представила уже пятую версию модели. Это свидетельствует о быстром развитии и изменении характеристик моделей и созданных ими текстов.

Таким образом, использование созданных нейросетями текстов для анализа языковой картины мира на данный момент не представляется методологически целесообразным.

Несмотря на кажущуюся иногда самостоятельность «мышления» нейросети, БЯМ – это продукт деятельности человека. И требования к модели закладываются разработчиками. Как правило, компании не публикуют полные данные об обучающих корпусах и требованиях к моделям, но в открытых документах можно найти сведения, указывающие на заложенные в модели установки. Это могут быть, например, политкорректность, многоязычность, кросс-культурная релевантность и т. п. Анализ требований, предъявляемых к большим языковым моделям, позволяет выявить, какие элементы *культурной картины мира* считаются важными в обществе, где разрабатываются и применяются эти технологии.

Понятие *культурной картины мира* включает в себя совокупность знаний, ценностей, норм, представлений и символов, которые характерны для определённого общества и отражаются в его языке, художественной и повседневной практике. В отличие от языковой картины мира, которая формируется в рамках конкретного языка, культурная картина мира шире: она охватывает как вербальные, так и невербальные способы интерпретации действительности [5; 6]. Если рассматривать феномен БЯМ через призму изучения языковой и культурной картины мира, кажется, что в настоящее время более целесообразным является анализ требований к нейросетям, нежели анализ сгенерированных текстов.

В современных исследованиях все чаще поднимается вопрос: какие культурные ценности и нормы транслируются через БЯМ? Анализ требований к моделям становится зеркалом культурных и этических ожиданий современного общества. Приведём примеры работ, посвященных исследованию кросс-культурной адаптивности моделей.

Группа ученых из Лондона (Рим И. Масуд, Цзыцюань Лю, Мартин Фериканк, Филип Треливен, Мигель Родригес) разработала собственный тест культурной согласованности ответов БЯМ (LLaMA 2, GPT-3.5 и GPT-4), основанный на шести культурных измерениях Хофстеде (дистанция власти, избегание неопределённости, индивидуализм/коллективизм, маскулинность/ фемининность, ориентация на долгосрочную/ краткосрочную перспективу, снисходительность/сдержанность). Ученые приходят к выводу, что из трех моделей GPT-4 лучше всех способна адаптироваться к различным культурным контекстам: особенно хорошо она показала себя в контексте китайской культуры, хуже справилась с США и арабскими странами. LLaMA 2, дообученная на китайском языке, продемонстрировала изменение в культурных паттернах ответов [7]. Таким образом, мы видим, что в более новой модели лучше проработана кросс-культурная адаптация ответов.

Американо-египетская команда ученых (Badr AlKhamissi, Muhammad ElNokrashy, Mai AlKhamissi, Mona Diab) оценивала культурное соответствие путем сравнения ответов моделей (GPT-3.5, AceGPT-Chat, LLaMA-2-Chat, mT0-XXL) с ответами реальных участников опроса в различных регионах Египта и Соединенных Штатов. Это исследование показало, что БЯМ демонстрируют большую культурную согласованность по двум направлениям: во-первых, при использовании доминирующего языка определенной культуры и, во-вторых, при предварительном обучении на языках, используемых в данной культуре. Также в работе подчеркивается необходимость создания более сбалансированного набора данных для подготовки к многоязычному обучению, чтобы лучше представлять разнообразие человеческого опыта и плюрализм культур [8].

Подведем итоги. Хотя большие языковые модели основаны на языке, использовать сгенерированные ими тексты для анализа языковой картины мира в полном объеме пока не представляется целесообразным из-за отсутствия устойчивой выборки вследствие большого количества текстов и нестабильности моделей. Более перспективным направлением представляется анализ требований, предъявляемых к таким моделям, как способ исследования современной культурной картины мира.

Список литературы:

1. Гумбольдт В. О различии строения человеческих языков и его влиянии на духовное развитие человечества // Избранные труды по языкознанию. М.: Прогресс, 1984. С. 34–298.

2. Сепир Э. Язык. Введение в изучение речи / пер. А.М. Сухотин. М.: Юрайт, 2025. 211 с.
3. Уорф Б.Ли. Язык, мышление, действительность / пер. с англ. В.И. Фролов; Московский государственный лингвистический университет. Москва: Издательская группа «Альма Матер», 2024. 333 с.
4. Нарумов Б. «Язык» лингвистики и «язык» философии. Contra Биbihин // Логос. 1999. № 1. С. 216.
5. Маслова В.А. Лингвокультурология: учебное пособие для вузов. М.: Академия, 2001. 204 с.
6. Степанов Ю.С. Константы. Словарь русской культуры. Опыт исследования. М.: Языки русской культуры, 1997. 824 с.
7. Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede's Cultural Dimensions / Reem I. Masoud, Ziquan Liu, Martin Feri-anc, Philip Treleaven, Miguel Rodrigues // Proceedings of the 31st International Conference on Computational Linguistics, January 19–24, 2025. 2025. P. 8474–8503. URL: <https://aclanthology.org/2025.coling-main.567.pdf> (дата обращения: 13.04.2025).
8. Investigating Cultural Alignment of Large Language Models / Badr AlKhamissi, Muhammad ElNokrashy, Mai AlKhamissi, Mona Diab // Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers), August 11–16, 2024. 2024. P. 12404–12422 URL: <https://aclanthology.org/2024.acl-long.671.pdf> (дата обращения: 13.04.2025).