

## Методологические риски интеграции генеративных моделей в современную лексикографическую практику

*Нурмухамедова Аннасолтан Аллабердыевна*

*Сотрудник*

Туркменский национальный институт мировых языков имени Довлетмаммета Азади,  
Ашхабад, Туркменистан  
*E-mail: temirka2016@gmail.com*

<р>Современная лексикография, находясь на стыке традиционной филологической науки и цифровых технологий, переживает этап глубокой методологической трансформации. Если в классических трудах Л. В. Щербы и В. В. Виноградова подчеркивалась необходимость строгого отбора языкового материала и системного отражения лексического строя, то в условиях цифровой эпохи к этим принципам добавляется новая переменная – искусственный интеллект (ИИ) как инструмент предварительной обработки и генерации данных. Актуальность исследования обусловлена динамичностью молодёжного сленга, при фиксации которого традиционные методы накопления материала оказываются недостаточно оперативными. В рамках проекта по формированию репрезентативного словника молодёжной речи был реализован полный цикл лексикографической работы с привлечением следующего технологического стека: • Генеративные модели: Использование ChatGPT (на базе GPT-4 и GPT-5) для оперативного получения списков актуальных лексем, сгруппированных по тематическим полям, с их предварительными толкованиями. • Корпусный инструментарий: Применение платформ Sketch Engine и AntConc для проверки контекстов реального функционирования слов, а также поисковых аналитических инструментов для оценки частотности и динамики употребления. • Автоматизированная обработка: Использование библиотек spaCy и morphology2 для морфологической обработки, лемматизации и последующей статистической фильтрации вариативных форм. Несмотря на технические преимущества (расширение охвата и ускорение рутинных операций), глубокий анализ выявил системные расхождения между результатами генерации ИИ и требованиями научной лексикографии: 1. Проблема лексических галлюцинаций. Вероятностный характер моделей приводит к созданию формально правдоподобных, но фактически фиктивных «псевдоединиц», не фиксируемых в реальных корпусах. Это ставит под сомнение принцип эмпирической проверяемости словника. 2. Искажение частотных характеристик. Алгоритмы цифровых платформ наследуют «перекося» интернет-корпусов, гиперрепрезентируя медийно тиражируемые тренды в ущерб устойчивым разговорным формам. В результате словник рискует отразить медиаповестку, а не живую практику. 3. Семантическая размытость. Черновые дефиниции ИИ часто носят избыточный характер и сглаживают прагматические оттенки – оценочную, ироническую и экспрессивную нагрузку, критически важную для сленга. Для нейтрализации данных рисков была внедрена система многоступенчатой валидации. Каждая единица проверялась в независимых источниках, а решение о включении принималось только при подтверждении устойчивого употребления экспертом. Все запросы к моделям фиксировались (версия, формулировка) для обеспечения воспроизводимости и научной прозрачности процедуры. Использование ИИ при составлении словников демонстрирует высокий потенциал для модернизации лексикографической практики, но требует строгого методологического контроля. Интеграция цифровых инструментов не отменяет фундаментальных принципов теории словаря, а дополняет профессиональную работу лексикографа, при условии, что окончательная интерпретация данных остается за специалистом. Дальнейшие перспективы связаны с созда-

нием специализированных обучающих корпусов для сленга и формализацией стандартов интеграции ИИ в академическую среду.</p>

#### **Источники и литература**

- 1) Ахманова, О.С. Словарь лингвистических терминов, М.: Либроком, 2013.
- 2) Малеева Д. А. Роль молодёжного и компьютерного сленга в системе языка // Филологические науки. 2011. №9. С. 19-21.
- 3) Милёшина, Л. В., Особенности молодёжного сленга в речи современного школьника // Балтийский гуманитарный журнал. 2021. №1. С. 365-368.
- 4) Щерба Л.В., Языковая система и речевая деятельность – Л.: Наука, 1974. – с. 266.