УДК 004.89:005.5

Л.С. ФАЙНЗІЛЬБЕРГ
Інститут інформаційних технологій та систем НАН України, Київ, Україна,
e-mail: *fainzilberg@gmail.com*.

ОБМЕЖЕННЯ АВТОРЕГРЕСІЙНИХ МОВНИХ МОДЕЛЕЙ У ЗАДАЧАХ ЛОГІЧНОГО ТА МАТЕМАТИЧНОГО ВИВЕДЕННЯ

Анотація. Проведено системний аналіз фундаментальних обмежень авторегресійних великих мовних моделей (ВММ), які використовуються в сучасних системах штучного інтелекту. На основі аналізу архітектури трансформерних моделей показано, що генерація тексту ґрунтується на статистичному прогнозуванні ймовірностей послідовностей токенів, що не забезпечує формалізованого логічного виведення. Окрему увагу приділено архітектурним обмеженням таких моделей у задачах математичного доведення, онтологічної інтерпретації та символічного міркування. Показано, що характерні помилки авторегресійних мовних моделей мають структурний характер і зумовлені архітектурними особливостями механізму генерації тексту. Запропоновано систематизацію архітектурних обмежень сучасних ВММ, яка може бути використана для аналізу перспективних мовних моделей та оцінювання їхньої здатності до логічного міркування. Отримані результати обґрунтовують доцільність інтеграції механізмів формального логічного виведення та онтологічного подання знань у перспективні архітектури штучного інтелекту.

Ключові слова: великі мовні моделі, системний аналіз, авторегресія, архітектурні обмеження, логічне міркування, формальна верифікація, онтологія, гібридні системи.

DOI 10.34229/KCA2522-9664.26.6.13

ВСТУП

Великі мовні моделі (ВММ) (Large Language Models, LLM), що належать класу систем штучного інтелекту, набули поширення завдяки здатності генерувати зв'язні та правдоподібні тексти в широкому спектрі задач [1–4]. Хоча відповіді моделей нерідко інтерпретуються як прояви інтелектуального міркування, така зовнішня подібність не свідчить про наявність у них стійкого семантичного або онтологічного представлення світу, а є наслідком особливостей архітектури та способу навчання авторегресійних моделей [5, 6].

Проблема архітектурних обмежень ВММ у задачах логічного та математичного виведення, що проявляється у схильності до галюцинацій та порушенні формальних правил, є об'єктом інтенсивних сучасних досліджень [7–9]. Зокрема, останні дослідження демонструють спроби подолання цих бар'єрів шляхом:

- розширення сприйняття моделі за межі тексту (мультимодальний контекст [10]);
- адаптації моделі до виконання нових задач безпосередньо під час діалогу за допомогою аналізу наданих прикладів (in-contextlearning [11]);
- динамічного перекодування контексту для підтримки логічної послідовності (реенкодинг контексту [12]).

Попри значний прогрес у розвитку архітектур ВММ та методів підвищення їхньої якості, сучасні підходи переважно спрямовані на зменшення частоти виникнення характерних помилок. Проте вони не усувають їхньої першопричини, оскільки не змінюють фундаментальний механізм авторегресійної генерації тексту, а здебільшого реалізуються як зовнішні процедури міркування, пошуку або верифікації. Можливість принципового усунення зазначених обмежень у межах наявної авторегресійної парадигми потребує подальшого дослідження [13–16].

Метою статті є аналіз особливостей архітектури сучасних авторегресійних ВММ та виявлення фундаментальних обмежень, зумовлених їхньою природою. Наукова новизна роботи полягає в обґрунтуванні того, що виявлені типи поми-

© Л.С. Файнзільберг, 2026