

УДК 004.8

Байдецька В.- ст. гр. КН-421

Відокремлений структурний підрозділ «Тернопільський фаховий коледж Тернопільського національного технічного університету імені Івана Пулюя»

ІМІТАЦІЯ СВІДОМОСТІ У ВЕЛИКИХ МОВНИХ МОДЕЛЯХ(LLM) ТА ПРИЧИНИ ЇЇ СПРИЙНЯТТЯ ЛЮДИНОЮ

Науковий керівник: викладач-методист, спеціаліст вищої категорії
Марціяш Г.Я.

Baydetska V.

Separate Structural Subdivision «Ternopil Professional College of Ternopil Ivan Pulyj National Technical University»

IMITATION OF CONSCIOUSNESS IN LARGE LANGUAGE MODELS AND REASONS FOR ITS HUMAN PERCEPTION

Supervisors: teacher-methodologist, specialist of the highest category
Martsiiyash G.

Ключові слова: свідомість, LLM, штучний інтелект, когнітивні процеси, ELIZA-ефект
Keywords: consciousness, LLM, artificial intelligence, cognitive processes, ELIZA effect

Здатність до творчості у людини базується на засвоєнні інформації з навколишнього середовища та її подальшій обробці. Аналогічно великі мовні моделі (LLM) навчаються на великих масивах даних для генерації нових текстових результатів. Попри це, людині традиційно приписується свідомість, тоді як штучні системи розглядаються як такі, що лише її імітують.

Свідомість є складним і багатовимірним поняттям, яке включає здатність до пізнання довкілля і самого себе через увагу, мислення, саморефлексію та метакогніцію. Важливим аспектом є суб'єктивний досвід (кваліа), який не піддається безпосередній верифікації ззовні. Це робить проблему визначення свідомості принципово відкритою.

Великі мовні моделі здатні ефективно імітувати окремі когнітивні процеси, зокрема мовлення, пам'ять і сприйняття, що створює ілюзію наявності розуміння.

Імітація мовлення у LLM базується на представленні слів у вигляді числових векторів та статистичному передбаченні найбільш ймовірного наступного елемента тексту. Попри складність обчислень, це не передбачає суб'єктивного розуміння змісту. Водночас людська поведінка також визначається багатьма чинниками, частина яких залишається невідомою, що ускладнює чітке розмежування "імітації" та "свідомості".

Імітація пам'яті у LLM базується на механізмах збереження та обробки даних. На відміну від людської пам'яті, яка є динамічним процесом (запам'ятовування, забування, реконструкція), комп'ютерна пам'ять є статичною та зберігає інформацію у незмінному вигляді до моменту її зміни або видалення.

Сприйняття також може бути імітоване на поведінковому рівні. Історичним прикладом є програма ELIZA, яка створювала ілюзію розуміння через перефразування відповідей користувача. Це призводило до емоційної залученості людей, попри

відсутність реального розуміння. Дане явище відоме як ELIZA-ефект і демонструє схильність людини до антропоморфізації.

Додатково сучасні дослідження компанії Anthropic показують, що внутрішні представлення мовних моделей можуть містити структуровані “вектори емоційних концептів”. За допомогою розріджених автоенкодерів (SAE) було виявлено 171 такий вектор у моделі Claude Sonnet 4.5, які впливають на поведінку моделі. Водночас ці структури не можуть бути ототоженні із суб’єктивним досвідом, але свідчать про складні внутрішні механізми обробки інформації.

Отже, великі мовні моделі не демонструють свідомості у філософському сенсі, однак їхня поведінкова складність та здатність до імітації когнітивних процесів формують у користувачів враження її наявності. Питання свідомості залишається відкритим і належить до міждисциплінарних проблем філософії, когнітивних наук та штучного інтелекту.

Список використаних джерел

1. Велитченко Л. К. Когнітивні процеси як суб’єктна динамічна система свідомості. 2011. УДК 159.9.
2. ELIZA effect – Wikipedia. URL: https://en.wikipedia.org/wiki/ELIZA_effect (дата звернення: 13.04.2026).
3. Anthropic. *Emotion Vectors in Large Language Models (Claude Sonnet 4.5)*, 2026. URL: <https://blog.pebbulous.ai/report/anthropic-emotions-report/en/>