

ПЕРЕВІРКА КОРЕКТНОСТІ ВІДПОВІДЕЙ АГЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ

Юрченко Василь Олександрович

здобувач магістерського рівня

Інститут атомної та теплової енергетики

Київський політехнічний інститут імені Ігоря Сікорського, Україна

Стрімке впровадження великих мовних моделей у прикладні галузі виявило ключову проблему — схильність ШІ до галюцинацій, тобто створення логічно або фактологічно хибних відповідей [1]. Цей виклик потребує впровадження механізмів автоматичної перевірки коректності. У роботі представлено багатоагентну систему CrossCheck AI, яка складається з трьох незалежних агентів: Responder, Verifier та Corrector. Кожен агент виконує власну роль, що дозволяє мінімізувати помилки та підвищити точність фінальних відповідей.

Для оцінювання ефективності системи було проведено серію експериментів із використанням різномірних моделей (GPT-4o [3], Claude 3 Opus [4], Gemini 1.5 Pro [5]).

Результати показали, що багатоагентний підхід значно зменшує кількість помилок порівняно з одноагентним режимом. На Рисунку 1 наведено порівняльний аналіз частоти помилкових відповідей у двох підходах. Як видно з графіка, використання CrossCheck AI знизило нормалізовану кількість помилок з 1.00 до 0.45, тобто більш ніж удвічі.

Зменшення кількості помилок після застосування CrossCheck AI

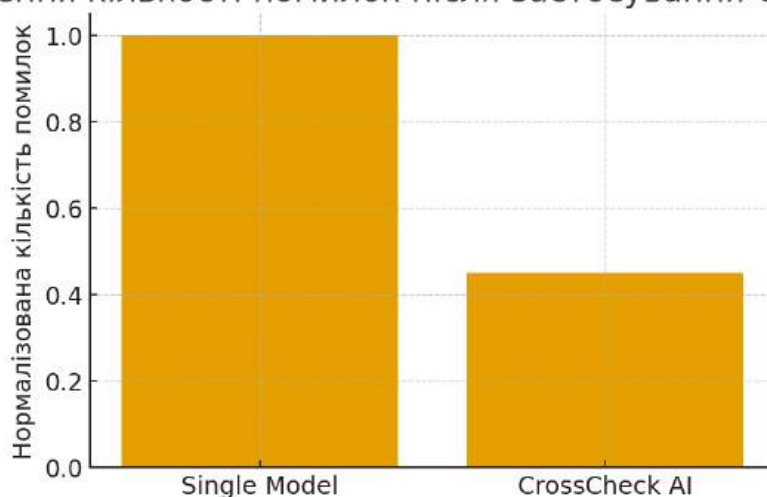


Рисунок 1 – Порівняння кількості помилок для одноагентного підходу та CrossCheck AI.

Додатковий аналіз показав, що застосування різних моделей у ролях окремих агентів суттєво підвищує точність відповіді. На Рисунку 2 наведено порівняння точності моделей у стандартному одноагентному режимі та при застосуванні CrossCheck AI.

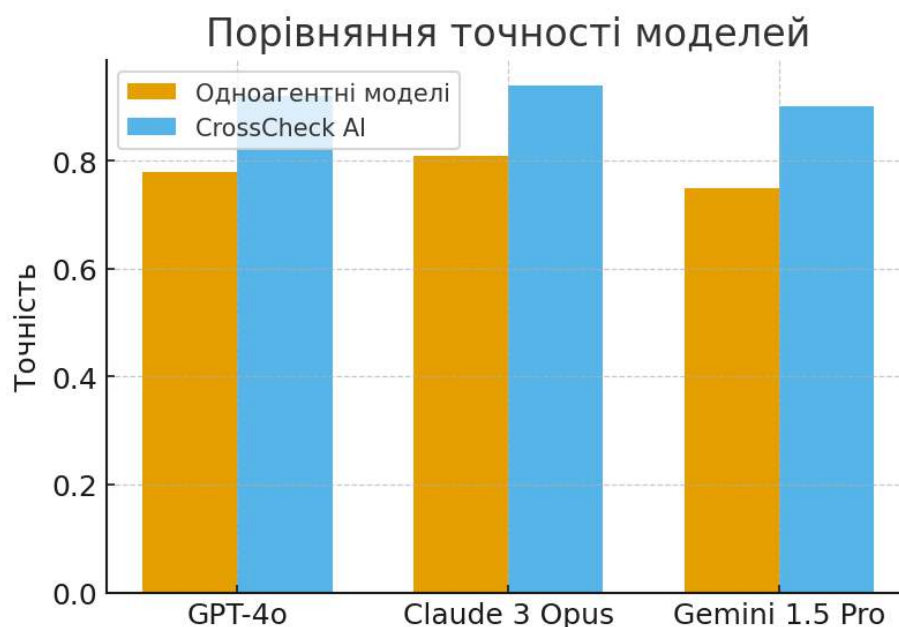


Рисунок 2 – Порівняння точності мовних моделей у різних режимах.

Точність GPT-4o зростає з 0.78 до 0.92 [3], Claude 3 Opus — з 0.81 до 0.94 [4], Gemini 1.5 Pro — з 0.75 до 0.90 [5]. Отримані дані узагальнено в Таблиці 1.

Таблиця 1 – Результати тестування точності CrossCheck AI

Модель	Точність (одноагентна)	Точність (CrossCheck AI)
GPT-4o	0.78	0.92
Claude 3 Opus	0.81	0.94
Gemini 1.5 Pro	0.75	0.90

Проведені експерименти також показали значне скорочення хибних відповідей навіть при роботі зі складними або багатокроковими запитами. Як видно з Таблиці 2, застосування CrossCheck AI зменшує кількість помилок майже вдвічі. Це дозволяє розглядати багатоагентні системи як один із найефективніших підходів боротьби з галюцинаціями ШІ.

Таблиця 2 – Зменшення кількості помилок після впровадження CrossCheck AI

Підхід	Помилки (нормалізовано)
Одноагентний	1.00
CrossCheck AI	0.45

Отримані результати підтверджують, що запропонована архітектура CrossCheck AI здатна покращити якість відповідей у критично важливих сферах — зокрема в аналітичних системах, енергетиці, технічній документації, системах підтримки прийняття рішень [1], [2]. Багатоагентний підхід може бути масштабований та адаптований для галузевих задач, де вартість помилки є високою, а надійність відповіді — визначальною.

Таким чином, результати досліджень демонструють, що CrossCheck AI є ефективним механізмом підвищення достовірності відповідей ШІ, забезпечуючи

стабільне зниження кількості помилок і значне зростання точності у порівнянні з традиційними одноагентними системами.

Список використаних джерел

1. Ji Z., Lee N., Frieske R. et al. Survey of Hallucination in Natural Language Generation. 2023.
2. Rawte V., Guo W., Shridhar K. Reducing LLM hallucinations using verifier models. 2023.
3. OpenAI. GPT-4 Technical Report. 2023.
4. Anthropic. Claude 3 Model Card. 2024.
5. Google DeepMind. Gemini 1.5 Research Overview. 2024.

DOI 10.70286/ISU-26.11.2025.006

PERSONALIZED RECOMMENDATIONS BASED ON THE PROCESSING OF TEXT DATA AND USER BEHAVIOR PATTERNS

Shapovalova Svitlana

Cand. Eng. Sc., Associate Professor

Huryh Iryna

Postgraduate student

Department of Digital Technologies in Energy

Igor Sikorsky Kyiv Polytechnic Institute, Ukraine

In modern information systems, personalized recommendations play an important role in improving the user experience when interacting with large amounts of digital content. Recommendation models are widely used in various services, which leads to a growing need for adaptive approaches to processing diverse information. This paper presents a comprehensive approach to creating a recommendation system based on the analysis of user interaction history and combining content and collaborative filtering methods. The Yelp Open Dataset, which includes both structured data about businesses and their interactions with users, as well as text reviews, was used to develop the system.

The main goal is to create a model that can generate relevant recommendations based on the user's previous experience. Three main approaches are used to achieve this goal: collaborative, content-based and hybrid filtering.

Collaborative filtering is based on identifying hidden factors in the matrix of interactions between users and businesses, which makes it possible to predict preferences based on analogies with other users [1]. The content approach uses semantic representations of businesses formed from text descriptions, categories, and attributes [2]. The hybrid approach combines content and collaborative models, complementing their strengths [3]. At the same time, its effectiveness directly depends