УДК 004.8

О. В. Бісікало
М. Ю. Дуднік
К. М. Даніленко

РОЗРОБКА ІНТЕЛЕКТУАЛЬНОГО ВЕБ-ЧАТУ З ТЕМАТИКИ ФОРМУЛИ-1 НА ОСНОВІ LLM

Вінницький національний технічний університет

Анотація

Розглянуто розробку інтелектуального веб-чату для отримання актуальної інформації з тематики Формули-1 із використанням великих мовних моделей. Запропонована система поєднує технології генерації з доповненою вибіркою, веб-пошуку та RSS-агрегації новин для формування відповідей у режимі реального часу. Реалізовано інтеграцію моделей Gemini 1.5 Flash та Llama 3.1 через API. Описано архітектуру програмного забезпечення, основні функціональні модулі та результати тестування системи.

Ключові слова: великі мовні моделі, веб-чат, Формула-1, генерація з доповненою вибіркою, штучний інтелект.

Abstract

The development of an intelligent web chat for obtaining up-to-date Formula-1 information using large language models is considered. The proposed system combines Retrieval-Augmented Generation technology, web search, and RSS news aggregation for generating responses in real time. Integration of the Gemini 1.5 Flash and Llama 3.1 models via API was implemented. The software architecture, main functional modules, and system testing results are described.

Keywords: large language models, web chat, Formula-1, Retrieval-Augmented Generation, artificial intelligence.

Вступ

Сучасні великі мовні моделі активно застосовуються для створення інтелектуальних інформаційних систем, здатних аналізувати природну мову та формувати відповіді, наближені до людського спілкування [1]. Одним із перспективних напрямів є створення спеціалізованих чат-асистентів, орієнтованих на конкретну предметну область. Розробка програмних засобів для реалізації таких прикладних задач є також важливою і з теоретичної точки зору. Актуальність таких засобів зростає, якщо вважати їх першим кроком до створення експериментальних майданчиків для удосконалення LLM на основі ідей та формалізмів теорії асоціативного образного мислення людини [2, 3].

Тематика Формули-1 характеризується великою кількістю динамічних даних: новини команд, результати перегонів, статистика пілотів, зміни регламенту та аналітичні матеріали. Через швидке оновлення інформації традиційні мовні моделі можуть надавати застарілі відповіді. Тому актуальною задачею є поєднання LLM із технологіями пошуку інформації в реальному часі [4].

Метою роботи є розробка інтелектуального веб-чату з тематики Формули-1, здатного автоматично знаходити, аналізувати та надавати актуальну інформацію користувачам.

Результати дослідження

Розроблений веб-чат реалізовано у вигляді веб-додатка із використанням мови програмування Python та фреймворку Streamlit [5]. Інтерфейс забезпечує взаємодію користувача з мовними моделями та відображення історії повідомлень.

Основою системи є технологія Retrieval-Augmented Generation, яка дозволяє поєднати генеративні можливості LLM із отриманням актуальних даних із зовнішніх джерел [4]. Для цього реалізовано модулі веб-пошуку та RSS-агрегації новин із тематичних ресурсів Формули-1.

Для забезпечення роботи системи використано дві великі мовні моделі:

1. Gemini 1.5 Flash – для глибокого аналізу великих обсягів інформації;
2. Llama 3.1 – для швидкої генерації відповідей із низькою затримкою.

Семантичний аналіз запитів дозволяє автоматично визначати категорію звернення користувача та обирати відповідне джерело інформації. Для пошуку актуальних новин реалізовано інтеграцію RSS-рядків спортивних ресурсів та веб-пошук через DuckDuckGo [6].

Інтерфейс розробленого веб-чату побудований таким чином, щоб забезпечити просту та зручну взаємодію користувача із системою. На головній сторінці користувач має можливість вводити текстові запити щодо новин Формули-1, статистики пілотів, результатів перегонів або інформації про команди. Приклад інтерфейсу програмної системи наведено на рис. 1.

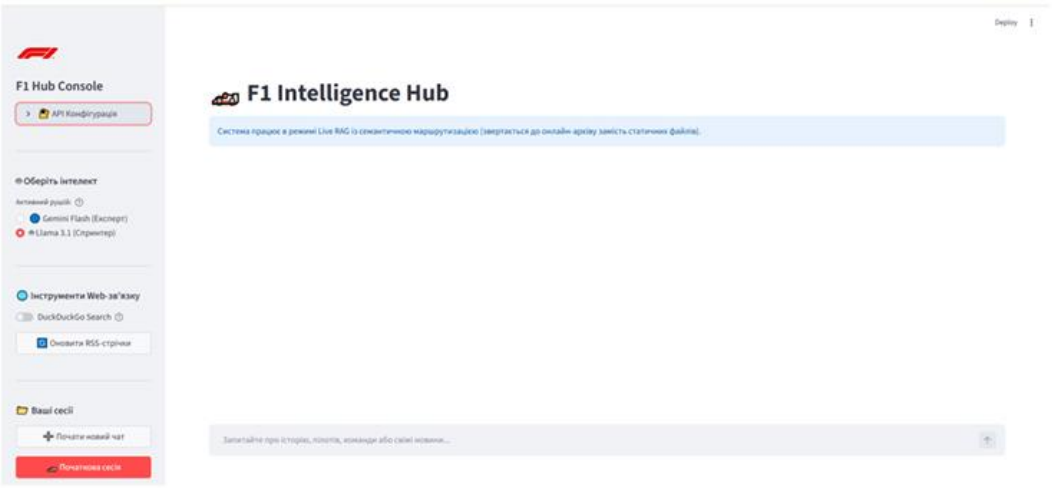

Рисунок 1 – Інтерфейс веб-чату

Після обробки запиту система автоматично визначає його тип та обирає відповідний механізм пошуку інформації. Для забезпечення актуальності відповідей реалізовано механізм Retrieval-Augmented Generation [4]. Під час отримання запиту система виконує пошук актуальної інформації через веб-пошук та RSS-стрічки новинних ресурсів. Отримані дані аналізуються та передаються мовній моделі як додатковий контекст для генерації відповіді. Приклад роботи веб-чату представлено на рис. 2.

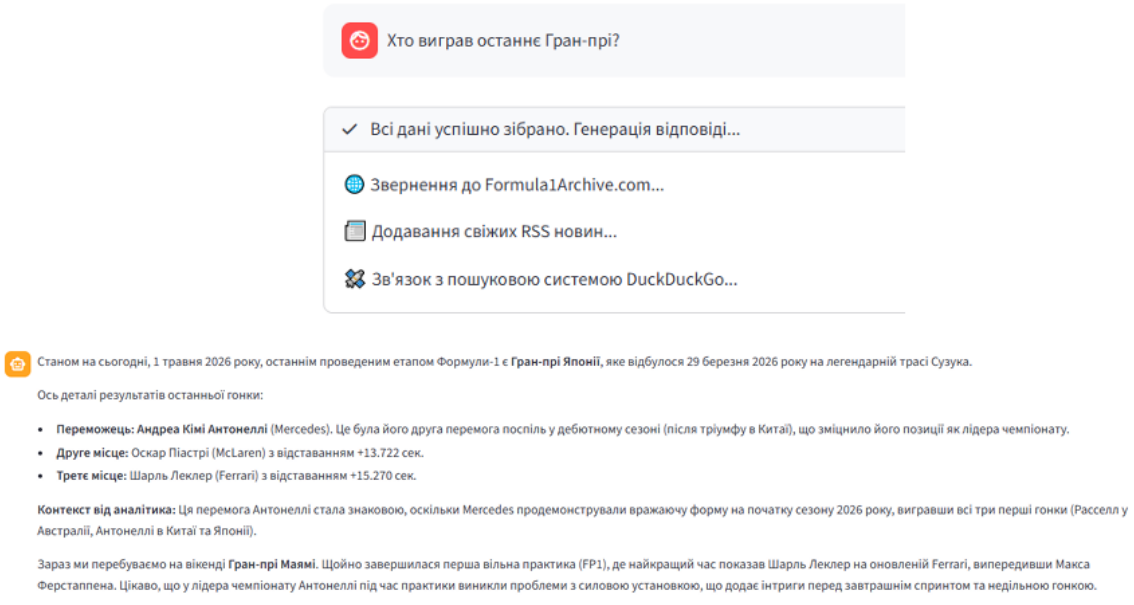

Рисунок 2 – Приклад роботи веб-чату

Архітектура системи включає декілька функціональних модулів:

1. модуль інтерфейсу користувача;
2. модуль семантичної маршрутизації запитів;
3. модулі веб-пошуку та обробки RSS-рядків;
4. модулі інтеграції великих мовних моделей;
5. модуль збереження історії чатів на базі SQLite [7].

Розроблений веб-додаток також підтримує збереження історії чатів у базі даних SQLite. Це дає змогу користувачу переглядати попередні повідомлення та повторно використовувати результати пошуку інформації. Крім того, реалізовано можливість оновлення RSS- рядків у режимі реального часу для автоматичного отримання свіжих новин зі світу Формули-1.

Висновки

У результаті роботи було розроблено інтелектуальний веб-чат для тематики Формули-1 на основі великих мовних моделей. Реалізована система забезпечує пошук та аналіз актуальної інформації в режимі реального часу, підтримує інтеграцію декількох LLM та дозволяє формувати інформативні відповіді користувачам.

Використання технології Retrieval-Augmented Generation дозволило поєднати можливості генеративного штучного інтелекту із механізмами веб-пошуку та RSS-агрегації новин. Це забезпечило підвищення актуальності та точності відповідей системи.

Результати тестування підтвердили ефективність використання моделей Gemini 1.5 Flash та Llama 3.1 для обробки тематичних запитів користувачів. Розроблений веб-додаток характеризується зручним інтерфейсом, швидкою обробкою запитів та можливістю подальшого розширення функціоналу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Brown T. Language Models are Few-Shot Learners / T. Brown, B. Mann, N. Ryder et al. // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 1877-1901.
2. Омельченко, В.; Бісікало, О. Удосконалення великих мовних моделей формальними засобами оцінки сенсу текстової інформації. ВНТКП ВНТУ. Факультет інтелектуальних інформаційних технологій та автоматизації, Ukraine, mar. 2026. Available at: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view/27435/22812>. Date accessed: 10 Mar. 2026.
3. Дадиверін В. В., Бісікало О. В. Аналіз та імплементація ідеї навчання великих мовних моделей за аналогією з дитячим когнітивним розвитком. – Вчені записки Таврійського національного університету імені В.І. Вернадського, Серія: Технічні науки, Том 36 (75) № 4 2025, Частина 2. – СС.103-109. – ISSN 2663-5941.
4. Lewis P. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks / P. Lewis, E. Perez, A. Piktus et al. // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 9459–9474.
5. Streamlit Documentation [Електронний ресурс]. – Режим доступу: https://streamlit.io/ (дата звернення: 10.05.2026).
6. SQLite Documentation [Електронний ресурс]. – Режим доступу: https://www.sqlite.org/docs.html (дата звернення: 10.05.2026).
7. DuckDuckGo Search Documentation [Електронний ресурс]. – Режим доступу: https://duckduckgo.com/ (дата звернення: 10.05.2026).

Бісікало Олег Володимирович – д-р техн. наук, професор, завідувач кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, email: obisikalo@vntu.edu.ua

Максим Юрійович Дуднік – студент групи 1ІСТ-23б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: maksimdudnik30@gmail.com

Даніленко Катерина Миколаївна – студентка групи 1ІСТ-24б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: viktoriadanilenko522@gmail.com

Bisikalo Oleh V. – Dr. Sc. (Eng.), Professor, Head of the Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University, Vinnytsia, email: obisikalo@vntu.edu.ua

Dudnik Maksym Yu. – student of group 1IST-23b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email: maksimdudnik30@gmail.com

Danilenko Kateryna M. – student of group 1IST-24b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email: viktoriadanilenko522@gmail.com