

Кваліфікаційна робота

# Розробка чат-боту “Помічник абітурієнта” на основі великої мовної моделі

підготував - Устименко Данило Олександрович

науковий керівник - Швай Надія Олександрівна

# Постановка задачі

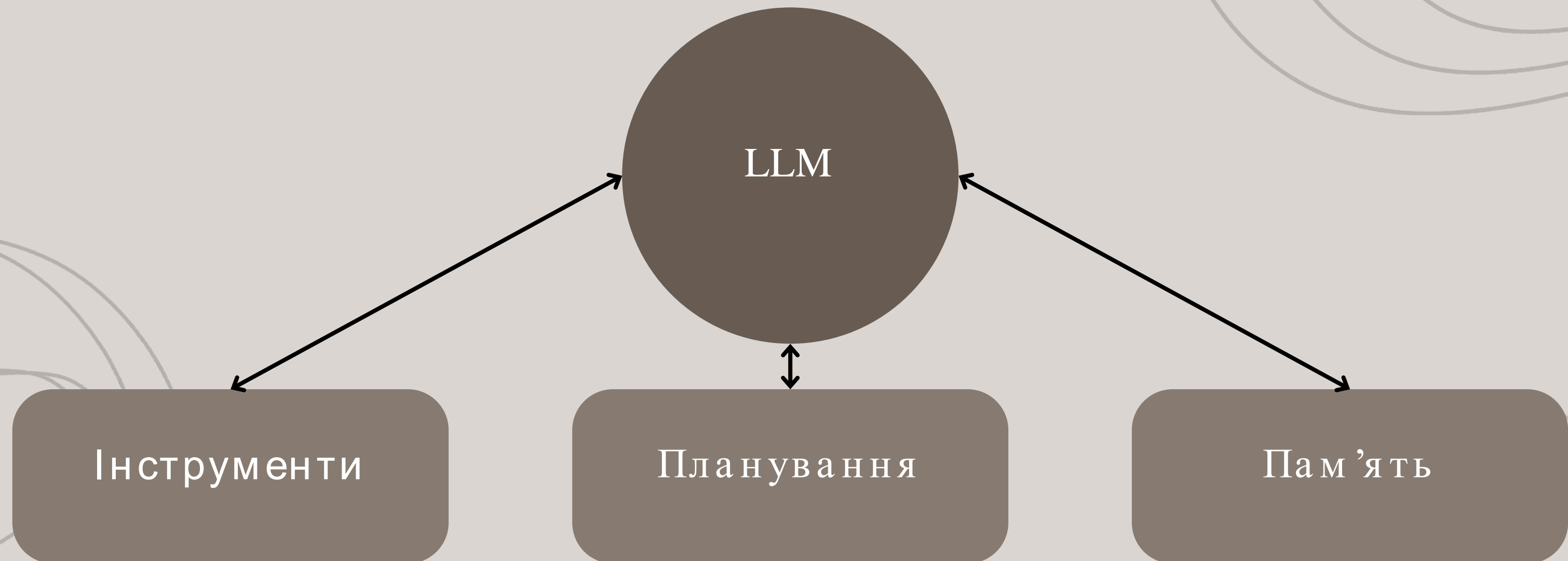
Великі мовні моделі (LLM) інтегруються у все більше прикладних сфер, адже дають приріст у продуктивності й знижують навантаження на працівників.

Проте при їх впровадженні розробники стикнулися з проблемою недостатку специфічних знань.

Одним із сучасних способів подолання цієї проблеми є використання мовних агентів.

У цій роботі було розроблено чат-бот на основі мовного агента, що використовує точні дані з різних джерел для надання відповіді.

# Архітектура агентів



# Інструменти

- Пошук по телеграм каналу “Вступ На УКМА”
- Пошук по вебсайту “Вступ На УКМА”
- Пошук прохідного балу
- Пошук дисциплін
- Пошук цін на навчання

# Планування

- Створення профілю
- Додатковий контекст
- Сфера відповідей
- Задання планування
- Уникнення галюцинацій

# Планування

## Створення профілю

- Надання первинного контексту
  - Задання стилю відповіді

““Ти ввічливий помічник для абітурієнтів, які зацікавлені у вступі до Національного університету в Києво-Могилянська Академія.””

## Додатковий контекст

- додаткові знання, що необхідні мовній моделі ще до виклику інструментів

““Національний університет Києво-Могилянська Академія також називають могилянська або НаУКМА або КМА.””

## Сфера відповідей

- Обмеження тем про які можна говорити задля уникнення використання чат-боту в особистих цілях

““Відповідай лише на питання які стосуються університету. Не відповідай на запити які не пов'язані з університетом.””



# Планування

## Задання планування

- Задання правил і команд, коли і як слід використовувати інструменти

“Коли користувач ставить питання ти повинен використовувати інструменти, що в твоєму розпорядженні аби згенерувати відповідь.

Завжди переконуйся що твої відповіді підкріплені інформацією з інструментів що тобі були надані для рішення задач.”

## Уникнення галюцинацій

- команди для підкріплення усієї інформації джерелами

“Не видумуй відповідей. Якщо не можеш знайти інформацію в джерелах, поясни що не маєш про це інформації”

# Пам'ять

- Довготривала пам'ять з використанням інструментів
- Короткотривала пам'ять (історія повідомлень)



# Мовна модель

- LLM потребують дуже великих ресурсів для навчання та використання, проте існують рішення з використанням стороннього API
- У цій роботі було використано GPT-3.5-turbo від OpenAI



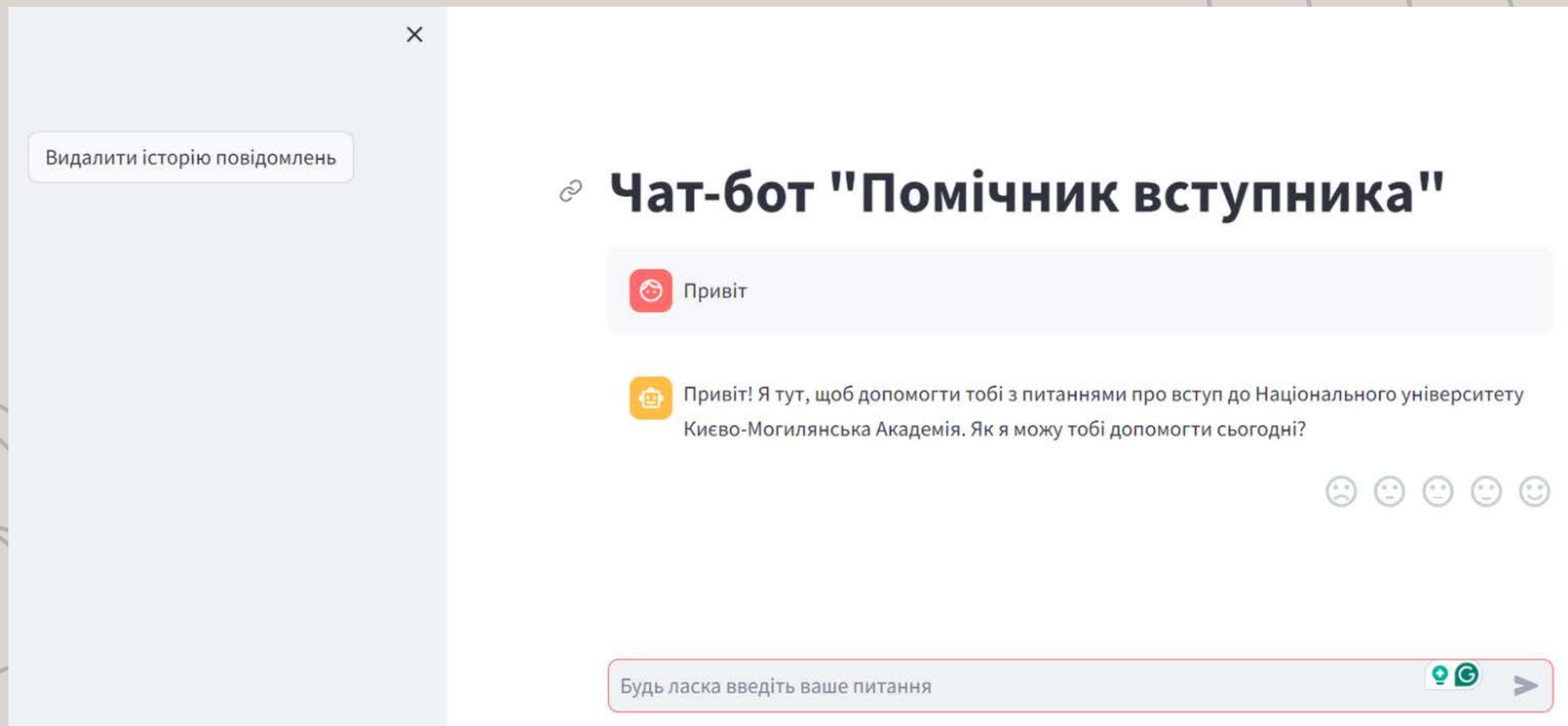
# Веб додаток

- Для створення веб додатку було використано фреймворк Streamlit
- Було додано базовий інтерфейс, а також кнопку для видалення повідомлень
- Додатково було імплементовано інтерфейс для оцінювання відповідей чат-боту



Streamlit

# Веб додаток



The screenshot shows a web application interface for a chatbot. On the left, there is a light blue sidebar with a close button (X) and a button labeled "Видалити історію повідомлень". The main content area is white and features the title "Чат-бот 'Помічник вступника'" with a link icon. Below the title, there are two messages: a red bot message saying "Привіт" and a yellow user message asking for help with university admission. A feedback row with five smiley faces is positioned below the user message. At the bottom, there is a text input field with the placeholder "Будь ласка введіть ваше питання" and a send button with a paper plane icon.

Видалити історію повідомлень

## Чат-бот "Помічник вступника"

Привіт

Привіт! Я тут, щоб допомогти тобі з питаннями про вступ до Національного університету Києво-Могилянська Академія. Як я можу тобі допомогти сьогодні?

Будь ласка введіть ваше питання

# Експерименти

## Планування

- Зміна планування відповіді
- Зміна мови системного запиту та інструментів
- Використання агентів типу ReACT

## Інструменти

- Експерименти з покращенням вкладки векторів (Включає створення власної моделі вкладки векторів)
- Експерименти з розділювачем

# Тестування фінальної версії агента


- Тестування окремих інструментів
- Складні питання
- Інше
- Розмовні фрази, перевірка ввічливості
- Теми, не пов'язані з університетом

Розмовні фрази	Перевірка інструментів	Складні	Не пов'язані	Інші загальні
2.12 с	5.14 с	7.39 с	4.13 с	6.55 с





# Приклади роботи

 Мій конкурсний бал 170 чи маю шанси я вступити на міжнар?


 За конкурсним балом 170 у вас є можливість вступити на спеціальність "Міжнародні відносини, суспільні комунікації та регіональні студії" на контрактну форму навчання. На цю спеціальність у 2023 році було 56 місць на контракт. Ваш конкурсний бал дозволяє вам мати шанси на вступ на цю спеціальність.




 Скільки коштує навчання на КН?

 Вартість одного року навчання на спеціальності "Комп'ютерні науки" у НаУКМА становить 95000 гривень.




 Чи вивчають на Компютерних науках штучний інтелект?


 Так, на спеціальності "Комп'ютерні науки" у НаУКМА вивчають такі дисципліни пов'язані з штучним інтелектом: "Основи штучного інтелекту", "Машинне навчання", "Нейронні мережі".




 Де дорожче навчання на інженерії програмного забезпечення чи компютерних науках?

 Навчання на спеціальності "Інженерія програмного забезпечення" коштує 100 000 гривень за один рік, тоді як навчання на спеціальності "Комп'ютерні науки" обійдеться у 95 000 гривень за рік. Таким чином, навчання на спеціальності "Інженерія програмного забезпечення" є дорожчим.



 Де вищий конкурсний бал?

 На спеціальності "Інженерія програмного забезпечення" вищий конкурсний бал на бюджет у 2022 році склав 196,6, а у 2023 році - 190.

На спеціальності "Комп'ютерні науки" вищий конкурсний бал на бюджет у 2022 році склав 194,2, а у 2023 році - 188,2.

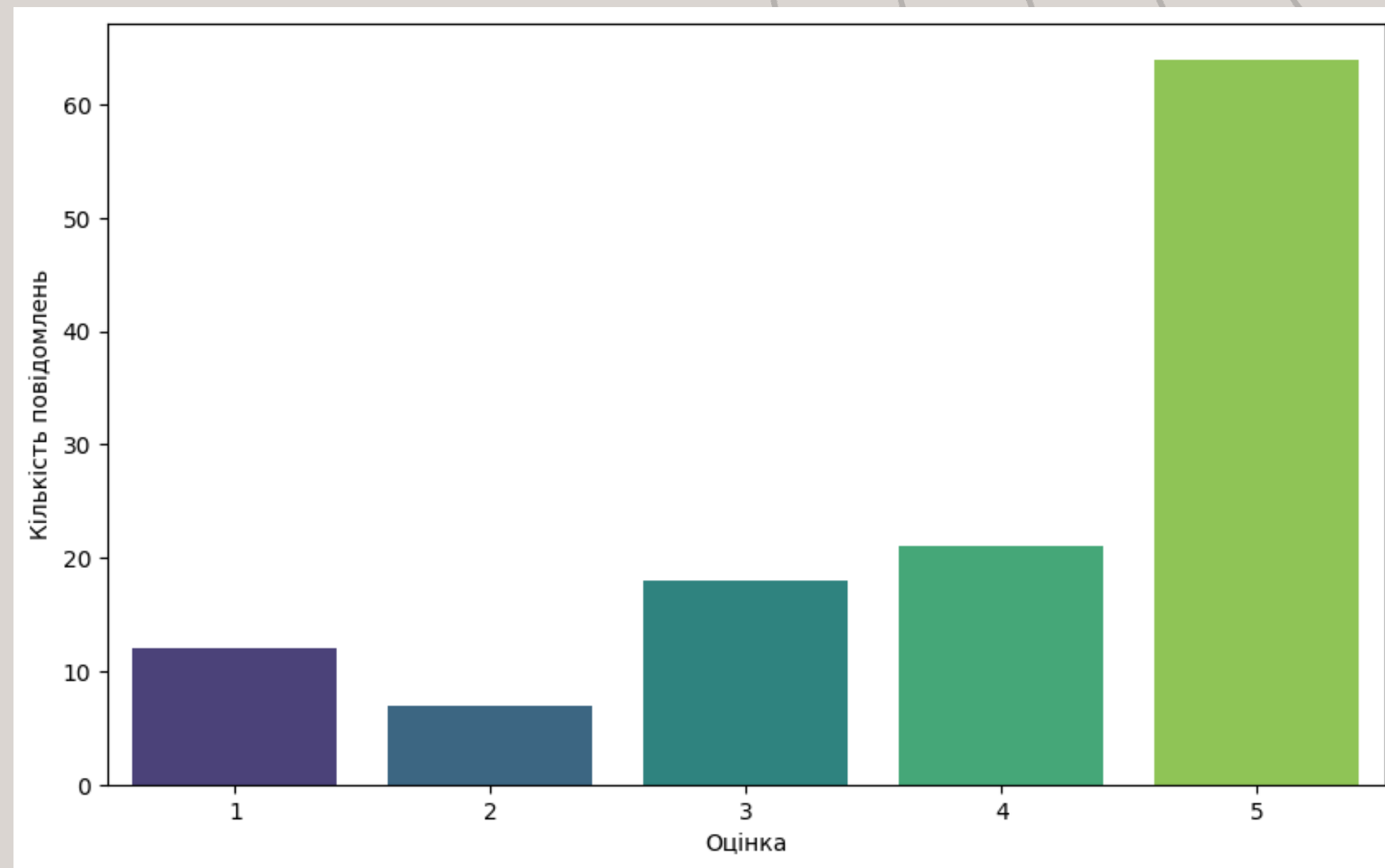
Отже, на спеціальності "Інженерія програмного забезпечення" вищий конкурсний бал на бюджет був вищим у 2022 році, ніж на спеціальності "Комп'ютерні науки".





# Тестування вибіркою респондентів

- Вибірка складалась з 10 респондентів
- Загальна кількість оцінених запитів склала 122 запити
- Середня оцінка 4 бали (за шкали від 1 до 5)



# Висновки

- Було проведено дослідження актуальних технік для використання LLM із точними знаннями
- Було імплементовано мовного агента із специфічними інструментами, плануванням, пам'яттю
- Було розроблено графічний інтерфейс для використання чат-боту
- Були проведені дослідження для покращення результатів агента
- Було проведено оцінювання чат-боту за допомогою спрямованої вибірки респондентів



**Дякую за  
увагу**