

**ІНТЕГРАЦІЯ RETRIEVAL - AUGMENTED GENERATION
У ПРОГРАМНУ СИСТЕМУ SOILSCOUT ДЛЯ ГЕНЕРАЦІЇ
ПЕРСОНАЛІЗОВАНИХ АГРОНОМІЧНИХ РЕКОМЕНДАЦІЙ**

Щербатюк М.О.

e-mail: maksym.shcherbatiuk@nure.ua

Науковий керівник – доц. каф. ПІ Лещинська І.О.

Харківський національний університет радіоелектроніки, каф. ПІ
м. Харків, Україна

This study explores enhancing the SoilScout IoT monitoring system by integrating artificial intelligence via the Retrieval-Augmented Generation approach. By linking large language models with 12 specialized datasets from Kaggle and Mendeley Data, the system provides personalized recommendations for crop selection and fertilization. The proposed architecture bridges the gap between real-time sensor data (pH, moisture, temperature) and expert agronomic knowledge. Implementing this AI module enables farmers to receive context-aware advice, significantly improving decision-making efficiency and crop yields. Future work focuses on predictive analytics and time-series data mining to prevent soil depletion and forestall critical changes in the soil.

Глобальна трансформація аграрного сектору в напрямку «сільського господарства 4.0» зумовлює необхідність впровадження інтелектуальних систем підтримки прийняття рішень [1]. Розроблена раніше система SoilScout успішно вирішує завдання оперативного збору даних про стан ґрунту, зокрема вологість, кислотність та температуру, за допомогою мобільного IoT-зонда. Однак наявність лише кількісних показників не завжди є достатньою для користувачів без спеціальної агрономічної підготовки, оскільки дані потребують професійної інтерпретації для перетворення їх на конкретні дії. Для перетворення системи моніторингу на повноцінного цифрового помічника виникає потреба в інтеграції компонентів штучного інтелекту, здатних аналізувати дані та надавати персоналізовані поради щодо вибору культур або норм внесення добрив.

Метою поточного дослідження є суттєве розширення функціональних можливостей системи SoilScout шляхом впровадження підходу RAG (Retrieval-Augmented Generation) [2]. Це дозволить використовувати великі мовні моделі (LLM) для генерації експертних агрономічних порад, спираючись не на загальні знання моделі, а на верифіковані спеціалізовані набори даних. Основним викликом при впровадженні штучного інтелекту в агротехнології є висока ціна помилки, оскільки стандартні моделі часто схильні до «галюцинацій» та надання узагальнених відповідей. Тому використання RAG-архітектури вирішує цю проблему шляхом створення зовнішньої векторної бази знань [3], пошуку

релевантного контексту за запитом користувача та формування відповіді мовної моделі з урахуванням виключно підтверджених фактів (див. рис. 1).

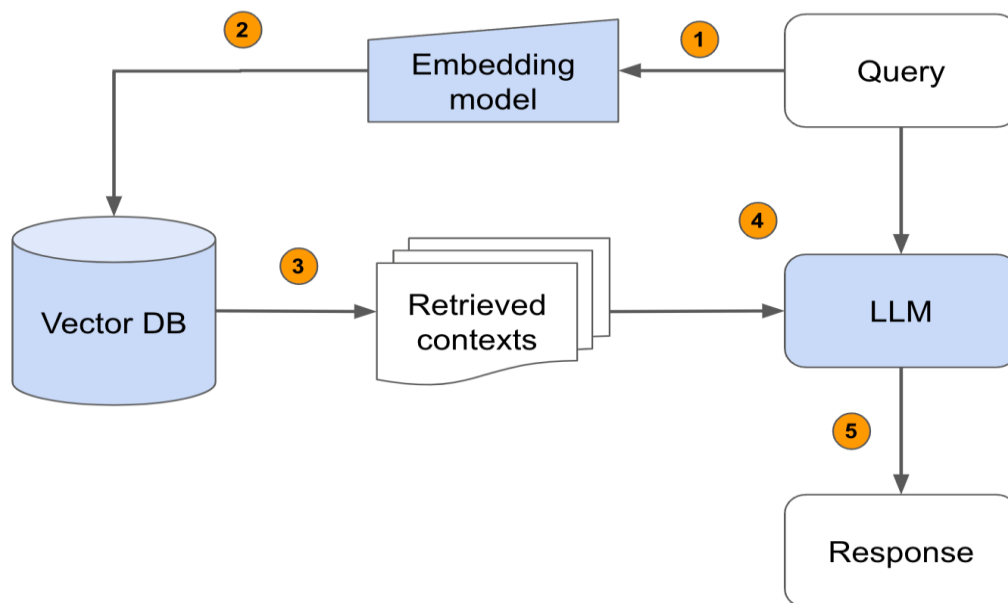


Рисунок 1 – Логічна схема функціонування RAG-модуля

Процес обробки даних у системі зав'язаний на моніторингу трьох фундаментальних показників, що визначають життєздатність рослин: кислотності, вологості та температури. Ключовим параметром є водневий показник (pH), який безпосередньо впливає на розчинність мінеральних сполук. Системою реалізовано класифікацію ґрунтів на сильнокислі ($\text{pH} < 4,5$), нейтральні ($6,0\text{--}7,2$) та лужні ($\text{pH} > 7,5$). Показники вологості інтерпретуються через призму водно-фізичних констант, зокрема критичних меж в'янення та польової вологості, що дозволяє формувати точні рекомендації щодо термінів та обсягів зрошення. Температурний режим ґрунту аналізується як фактор, що регулює швидкість поглинання вологи кореневою системою та активність ґрунтової мікрофлори.

Для наповнення бази знань було проаналізовано та відібрано 12 ключових датасетів з платформ Kaggle та Mendeley Data [4]. Зокрема, набори даних «Crop and Soil Dataset» та «Plants Growth and Care Recommendations» дозволяють встановити чіткі кореляції між рівнем pH ґрунту та оптимальними фазами росту рослин. Датасети, що містять інформацію про врожайність та екологічні фактори за 2014–2023 роки, використовуються для прогнозування економічної ефективності вирощування конкретних культур у заданих кліматичних умовах. Важливою частиною роботи є аналіз спеціалізованих даних щодо внесення добрив, таких як «Fertilizer Recommendation Dataset», що дає змогу

інтегрувати в систему алгоритми розрахунку мінеральних підживлень залежно від хімічного складу ґрунту.

Програмна реалізація оновленої системи включає розширення серверної частини на базі Node.js шляхом впровадження окремого мікросервісу для координації запитів між мовною моделлю та векторним сховищем знань. У межах модернізації структури даних було доповнено базу MySQL новими атрибутами, зокрема типом ґрунту та видом сільськогосподарської культури, що вирощується на конкретній ділянці. Це дозволяє алгоритмам штучного інтелекту враховувати індивідуальні особливості поля при аналізі історії вимірювань. В оновленому інтерфейсі веб-платформи та мобільного застосунку реалізовано розширені функціональні можливості для інтеграції спеціалізованого блоку візуалізації генеративних порад.

Впровадження RAG-підходу дозволило якісно розширити можливості системи SoilScout. Завдяки використанню зовнішніх датасетів очікується досягнення значно вищої точності рекомендацій, ніж при використанні базових моделей без додаткового контексту. Це мінімізує ризики неправильного догляду за посівами та оптимізує витрати на ресурси. Подальший розвиток проєкту передбачає впровадження методів Data Mining для аналізу часових рядів вимірювань. Це дозволить прогнозувати динаміку виснаження ґрунту та надавати превентивні поради ще до моменту настання критичних змін, забезпечуючи сталий розвиток господарства та збереження родючості земель у довгостроковій перспективі.

Список використаних джерел:

1. Служба новин. Сільське господарство 4.0 в Україні. Цифрові технології на практиці. Куркуль – онлайн-асистент фермера. URL: <https://kurkul.com/spetsproekty/1743-silске-gospodarstvo-40--priklad-rozvitku-v-ukrayini> (дата звернення: 12.03.2026).
2. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv.org. URL: <https://arxiv.org/abs/2005.11401> (дата звернення: 11.03.2026).
3. Roie Schwaber-Cohen. What is a Vector Database & How Does it Work? Use Cases + Examples | Pinecone. The vector database to build knowledgeable AI | Pinecone. URL: <https://www.pinecone.io/learn/vector-database/> (дата звернення: 12.03.2026).
4. Alemu S. Crop Recommendation using Soil Properties and Weather Prediction Dataset. Mendeley Data. URL: <https://data.mendeley.com/datasets/8v757rr4st/1> (дата звернення: 11.03.2026).