

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ В. Н. КАРАЗІНА

ІТ-ПРОСТІР СЬОГОДЕННЯ: ТЕНДЕНЦІЇ, ІННОВАЦІЇ ТА ПЕРСПЕКТИВИ РОЗВИТКУ

Збірник тез доповідей
II Міжнародної науково-практичної студентської
конференції

(15 жовтня 2025 року, м. Харків, Україна)

Електронний ресурс

Харків – 2025

Астаф'єва К.Ю.	
ВИКЛИКИ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ В УКРАЇНСЬКОМУ БІЗНЕСІ ТА МОЖЛИВІ РІШЕННЯ.....	318
Білоцерківська В.А.	
МЕТОДИ ЗБОРУ ТА ПОПЕРЕДНЬОЇ ОБРОБКИ ДАНИХ ВАКАНСІЙ ДЛЯ ПОДАЛЬШОГО МАШИННОГО АНАЛІЗУ	320
Бобейко К.С., Кобилін О.А.	
ІНТЕЛЕКТУАЛЬНІ ПРОГРАМНІ ЗАСОБИ ДЛЯ ЕНЕРГОЗБЕРЕЖЕННЯ В ЖИТЛОВОМУ СЕКТОРІ.....	322
Богдан Н.І.	
ПОРІВНЯЛЬНИЙ АНАЛІЗ АРХІТЕКТУРНИХ РІШЕНЬ ДЛЯ ІНТЕГРАЦІЇ ДОВГОТРИВАЛОЇ ПАМ'ЯТІ У ДІАЛОГОВІ СИСТЕМИ.....	325
Бойко М.К.	
МЕТОД БАГАТОВИМІРНОГО АНАЛІЗУ ЕФЕКТИВНОСТІ РЕКЛАМНИХ КАМПАНІЙ З АДАПТИВНИМ МОНІТОРИНГОМ.....	327
Бондаренко А. Є.	
АНАЛІЗ МІЖНАРОДНИХ РИНКІВ ДЛЯ ОПТИМІЗАЦІЇ ПОСТАЧАННЯ.....	329
Будник В.І.	
ПЕРСПЕКТИВИ ВИКОРИСТАННЯ БЛОКЧЕЙН-ТЕХНОЛОГІЙ В ОБЛІКУ ЦИФРОВИХ АКТИВІВ.....	332
Веприцький А.В.	
СИСТЕМИ БІЗНЕС-АНАЛІТИКИ ТА МАШИННЕ НАВЧАННЯ: СУЧАСНІ РІШЕННЯ ДЛЯ ПІДТРИМКИ ПРИЙНЯТТЯ УПРАВЛІНСЬКИХ РІШЕНЬ.....	335
Вигоднер І.В., Полозова С.Є.	
РОЗВ'ЯЗУВАННЯ ОПТИМІЗАЦІЙНИХ ЗАДАЧ ЗА ДОПОМОГОЮ ТАБЛИЧНОГО ПРОЦЕСОРУ EXCEL.....	337
Власов І.В., Томашук І.В.	
ІННОВАЦІЇ У РОЗВИТКУ МАЛОГО ТА СЕРЕДНЬОГО БІЗНЕСУ ЯК ФАКТОР ЕКОНОМІЧНОГО ЗРОСТАННЯ.....	340
Галчанська В.В.	
ПЕРЕВАГИ ІНФОРМАЦІЙНИХ СИСТЕМ В КОНТЕКСТІ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ СУСПІЛЬСТВА.....	343
Глушко Р. О.	
ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА РОЗВИТОК ФІНАНСОВИХ ТЕХНОЛОГІЙ.....	345
Гончарук Д.О., Заїченко Ю.П.	
ОСВІТНІ ПЛАТФОРМИ ТА ЗАСТОСУНКИ НОВОГО ПОКОЛІННЯ.....	348

Богдан Н.І.
здобувач вищої освіти,
Харківський національний університет радіоелектроніки
Науковий керівник
Творошенко І.С.
к.т.н., доцент, доцент кафедри інформатики
Харківський національний університет радіоелектроніки

ПОРІВНЯЛЬНИЙ АНАЛІЗ АРХІТЕКТУРНИХ РІШЕНЬ ДЛЯ ІНТЕГРАЦІЇ ДОВГОТРИВАЛОЇ ПАМ'ЯТІ У ДІАЛогові СИСТЕМИ

Тематика обмеженої контекстуальної пам'яті існуючих великих мовних моделей (LLMs) є одним із ключових архітектурних викликів у проєктуванні сучасних діалогових систем. Кінцевий розмір контекстного вікна призводить до втрати релевантної інформації під час тривалих сесій, що знижує актуальність та якість взаємодії. Це, в свою чергу, вимагає пошуку та аналізу архітектурних рішень, здатних розширити пам'ять LLMs, забезпечуючи послідовність і логічну узгодженість діалогу [1-3].

Для подолання вказаної проблеми існує низка підходів, що принципово відрізняються своєю архітектурою та ефективністю:

- архітектура «ковзного вікна» (Sliding Window) – метод, який є найпростішим з точки зору реалізації. Він передбачає збереження лише останніх повідомлень діалогу, які вміщуються у фіксований буфер контекстного вікна. Ключовою перевагою даного методу є простота і низькі обчислювальні витрати. Однак, зазначений підхід має суттєвий недолік: він безповоротно втрачає ранній контекст, що робить його непридатним для тривалих діалогів, де важлива вся історія;

- архітектура узагальненої (епізодичної) пам'яті (Summarized/Episodic Memory) – метод, який є проміжним між простим буфером і складними гібридними системами. Його архітектура включає додатковий етап, на якому старі фрагменти діалогу періодично узагальнюються в короткі резюме. Ці резюме потім додаються до поточного контекстного вікна. Перевага цієї архітектури полягає в збереженні ключових ідей розмови без перевантаження контексту, вона може втрачати дрібні, але важливі деталі, які не потрапили в узагальнення;

- архітектура «генерації, доповненої пошуком» (Retrieval-Augmented Generation, RAG) – найбільш ефективне архітектурне рішення. Метод поєднує можливості LLM з зовнішньою базою знань. Архітектура RAG складається із двох основних компонентів: ретрівер-модуля, що виконує пошук релевантних фрагментів даних, і генератор-модуля, який створює відповідь, використовуючи як запит, так і знайдені дані (рис. 1). Цей підхід забезпечує високу точність, дозволяє надавати відповіді, які ґрунтуються на достовірних джерелах, і значно розширює «пам'ять» системи. Однак, його головним недоліком є архітектурна

складність та потреба в додаткових ресурсах для керування зовнішньою базою даних.

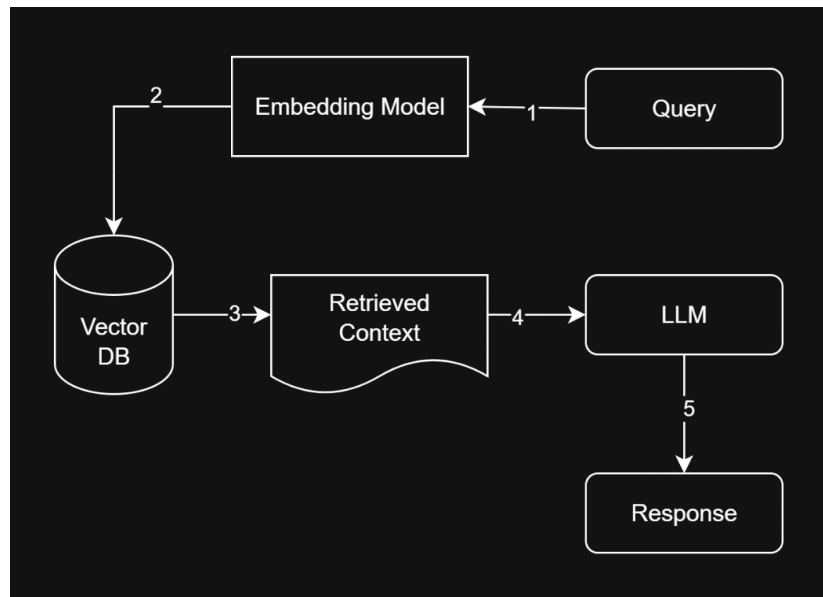


Рисунок 1 – Схематичне подання методу RAG

Отже, вибір архітектурного рішення для інтеграції довготривалої пам'яті є компромісом між простотою реалізації та ефективністю. Якщо «ковзне вікно» підходить для простих завдань, то для критично важливих застосувань, що вимагають високої точності та зв'язності, архітектура RAG є найбільш доцільною. Розуміння цих архітектурних відмінностей є фундаментальним для проектування та впровадження надійних і ефективних діалогових систем [4, 5].

Список використаних джерел:

1. Gorokhovatskyi V., Chmutov Y., Tvoroshenko I., and Kobylin O. (2025) Reducing computational costs by compressing the structural description in image classification methods, *Advanced Information Systems*, vol. 9, no. 1, pp. 5-12.
2. Gorokhovatskyi V., Tvoroshenko I., Yakovleva O., Hudáková M., and Gorokhovatskyi O. (2024) Application a committee of Kohonen neural networks to training of image classifier based on description of descriptors set, *IEEE Access*, vol. 12, pp. 73376-73385.
3. Gorokhovatskyi V., Tvoroshenko I., Yakovleva O., and Hudáková M. (2025) Image description compression in classification structural methods, *IEEE Access*, vol. 13, pp. 43631-43641.
4. Yakovleva O., Matúšová S., Tvoroshenko I., and Isaiev Y. (2024) Visitor counting based on video stream analysis from surveillance cameras to solve various business problems, *Verejná správa a regionálny rozvoj ekonómia, manažment a marketing*, XX(1), pp. 67-87.
5. Gorokhovatskyi V., Tvoroshenko I., Yakovleva O. (2024) Transforming image descriptions as a set of descriptors to construct classification features, *Indonesian Journal of Electrical Engineering and Computer Science*, 33 (1), 113-125.