

## НАПРЯМ 1. ЕКОНОМІКА

**Гром'як С.І.**

*кандидат економічних наук, доцент,  
докторант,*

*Львівський національний університет ветеринарної медицини  
та біотехнологій імені С.З. Гжицького*

DOI: <https://doi.org/10.36059/978-966-397-482-8-1>

### **ЗМАГАЛЬНИЙ МЕТОД ВИКОРИСТАННЯ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ВИБОРУ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ**

Системи підтримки прийняття рішень, далі по тексті СППР, спеціалізуються на вирішенні широкого кола управлінських задач, однією з яких є задача вибору. У найпростішому варіанті – це задача вибору серед *визначеного кола варіантів*, яку вирішують методом ранжування, або вибору найкращого варіанту за набором кількох критерієм, або ж методом експертних оцінок з використанням теорії нечітких множин, а також каскадний (багатоетапний) відбір. Більш складними варіантами задачі є так звана *задача вередливої нареченої та теорія ігор з нульовою сумою виграшу*, обидві з обмеженим колом варіантів, але обмеженням по часу та ресурсах, які затрачаються на вибір. Окремим варіантом є задача з *невизначеним числом варіантів вибору*. Такі задачі, зокрема, часто зустрічаються при попередньому виборі варіантів архітектурних рішень для інформаційних систем. Результатом вирішення задачі є документ «картка вибору архітектурного рішення».

Розглянемо приклад: потрібно підібрати архітектурне рішення для генерації звіту на основі десятків мільйонів однотипних документів (рахунків за послуги), які генеруються щомісячно. Вказані вхідні дані (йдеться про кожен окремий рахунок) представлені ієрархічною структурою без нормалізації відношень.

Архітектурної групою визначено, що головними критеріями будуть мінімізація затрат на обробку та швидкість формування звіту, очевидно (прим. авт.: мається на увазі, що це є очевидним для досвідчених архітекторів ІС), що потрібно буде вибрати одне з рішень BI (business intelligence) або ж одну з OLAP баз даних, іншою альтернативою є адаптувати одну з існуючих компонент інформаційної системи (базу даних), які вже використовуються, для вирішення задачі генерації звітів.

Зауважмо, що вже на етапі постановки задачі маємо: а) нечіткі критерії вибору, б) негомогенні типи варіантів рішення задачі (BI, OLAP, традиційні бази даних та алгоритмічні методи), та в) немає уявлення про загальну кількість варіантів для вибору.

Якщо взяти за основу традиційний рейтинг баз даних то лише для OLAP технології у нас буде більше кількох десятків варіантів.

Для вирішення подібного типу нечітких задач зазвичай використовується метод експертних оцінок, який базується на досвіді експертів та на математичному апараті теорії нечітких множин [3]. СППР при цьому виступають у ролі інструмента, який дозволяє генералізувати такі оцінки і оркеструє процес прийняття рішення.

Іншим способом є *евристичний аналіз* неструктурованої інформації, як-от: 1) опису постановки задачі, 2) опису існуючої системи та 3) інформації з відкритих джерел про наявні архітектурні рішення. Для цього можна використати великі мовні моделі (LLM), далі по тексту мовні моделі.

При використанні мовних моделей у нас з'являється додаткове нефункціональне обмеження: захист інформації у процесі розв'язування задачі. Це тому, що частина вхідної інформації (постановка задачі, опис системи), як правило, має обмежений доступ у межах організації, де потрібно впровадити рішення.

Нефункціональна частина пов'язана з доступом до інформації може бути вирішена двома способами: використання комерційних мовних моделей для корпоративних користувачів, таких як Azure Services від Microsoft. У цьому варіанті маємо обмежене коло варіантів вибору моделі для аналізу. Інший

варіант – це запуск невеликої моделі локально для доступу до приватних даних та анонімізації запиту до моделі з публічним доступом (така обробка може бути реалізована обов'язковою моделі, тобто спеціально розробленим середовищем для запуску мовних моделей на основі LangGraph, LangChain, n8n, тощо). Обидва варіанти достатньо описані у [1].

Отже, для ефективного вирішення задачі мовними моделями потрібно провести її декомпозицію. Щодо декомпозиції є два варіанти:

**Варіант 1.** Можна делегувати цей етап для мовної моделі, в яку вбудована функція розважання (*reasoning*) або глибоке дослідження (*deep research*) [2]. Але для цього етапу необхідний доступ до інформації з обмеженим доступом. У випадку, коли використовується локально запущена модель, то її розміри і функціонал доволі обмежені, тобто не завжди є можливість використати розважання.

**Варіант 2.** Декомпозицію може зробити людина-експерт. Отже, вже на цьому етапі виникаю вимоги до компетенції, досвіду і навиків людини експерта. Зважаючи, що мовні моделі перебувають лише на другому етапі свого розвитку за С. Альтманом – це етап асистента, то в цьому немає нічого дивного.

Проведемо декомпозицію задачі, розбивши на декілька етапів:

- 1) розбивка на етапи (власне це вона і є);
- 2) визначення вбивчих факторів (*killing factors*) та критеріїв оцінки вибору;
- 3) анонімізація внутрішніх варіантів вибору (для окреслення яких необхідна інформація з обмеженим доступом);
- 4) попередній аналіз типів варіантів вибору;
- 5) підбір варіантів рішення;
- 6) порівняльний аналіз рішень;
- 7) узагальнення вибору;
- 8) оцінка вибраного варіанту.

Якщо використовуються мовні моделі для корпоративного середовища, то вже на першому ж етапі можна використати змагальне використання мовних моделей.

При змагальному методі завдання виконується на двох мовних моделях паралельно, результат трактується в залежності від обраного варіанту використання. Отже є два варіанти використання:

**Варіант 1.** Беруться мала і велика мовні моделі. Спершу виконується контрольне завдання (в рамках поставленої задачі) на малій моделі, якщо «результат перевершує всі очікування», то більша модель не застосовується, інакше робимо виконання на більшій моделі, якщо результат малої моделі не критично гірший, виконання продовжується на малій, інакше на великій або на обох одразу. Такий варіант найкраще підходить для експрес-аналізу.

**Варіант 2.** Завдання виконується на кількох моделях паралельно. При цьому якщо це експрес-аналіз, вибирається результат який найкраще співпадає з оцінками експерта по ключових характеристиках. Якщо ж це повний, або поглиблений аналіз, то кількісні результати моделі трактуються як функція належності й генералізуються з використанням математичного апарату нечітких множин. Для другого варіанту обирається до 10 моделей (у випадку якщо треба використати ресурс на зразок `lmarena.ai` [4], де попарно запускаються одні з найкращих моделей. Для цього варіанту запити будуть анонімізовані, тобто очищені від РІІ або іншої інформації з обмеженим доступом).

Для ефективного виконання задач за допомогою мовних моделей потрібно провести якісну декомпозицію задачі.

На сьогоднішній день кваліфікація експерта відіграє ключову роль у вирішенні задач з допомогою мовних моделей.

Мовні моделі дозволяють вирішувати задачі на неструктурованих та слабо-структурованих даних та з використанням бази знань.

### Література:

1. Громяк С.І. Адаптація великих мовних моделей для вузькоспеціалізованих предметних областей на прикладі аграрного сектору. *Наукові перспективи*. 2025. № 1 (55). С. 765–788.
2. Brodsky S. OpenAI's deep research aims to outthink analysts. IBM Think. URL: <https://www.ibm.com/think/news/openai-releases-deep-research> (дата звернення: 06.03.2025).
3. Zadeh L. A. Fuzzy sets. *Information and Control*. 1965. No. 8(3). P. 338–353.
4. Chiang, Wei-Lin, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, et al. Chatbot Arena: An Open

Platform for Evaluating LLMs by Human Preference. arXiv. 07.03.2024.  
DOI: <https://doi.org/10.48550/arXiv.2403.04132> (дата звернення:  
06.03.2025).

5. Козлов О. В., Кондратенко Ю. П. Методи та моделі інтелектуальних обчислень : навчальний посібник. Миколаїв : ЧНУ, 2024. 148 с.