Міжнародний конкурс студентських робіт зі штучного інтелекту 2026

II тур
Перший етап

Р Е Ц Е Н З І Я

на наукову роботу

назва Метод контекстно-залежного пошуку в ієрархічних RAG-системах зі збереженням сутностей

шифр ГРАНУЛЯРНІСТЬ-RAG

представлену на Міжнародний конкурс студентських робіт зі штучного інтелекту

№ з/п	Характеристики та критерії оцінки рукопису наукової роботи	Рейтингова оцінка. Максимальна кількість балів (за 120-бальною шкалою)	Бали
1	Актуальність проблеми	10	10,0
2	Новизна та оригінальність ідей	15	11,8
3	Використані методи дослідження	15	12,0
4	Теоретичні наукові результати	10	8,5
5	Практична направленість результатів	20	14,8
6	Рівень використання наукової літератури та інших джерел інформації	5	5,0
7	Ступінь самостійності роботи	10	8,8
8	Якість оформлення	5	4,3
9	Наукові публікації	10	0,0
10	Недоліки роботи (пояснення зниження максимальних балів у пунктах 1-9):		

Дослідження добре вмотивоване, має чітку постановку, продуману абляційну схему та статистичну обробку результатів, що вигідно вирізняє його серед робіт секції.

Запропоноване поєднання збереження сутностей, атрибутів і відношень із маршрутизацією за рівнями ієрархії має дослідницьку цінність. Водночас оригінальність окремих складових та їх відмінності від найближчих методів структурованої сумаризації й адаптивного пошуку потребують чіткішого порівняння.

Тестова вибірка охоплює лише 40 запитів, спеціально відібраних за наявністю числових параметрів і потребою міждокументного синтезу. Це відповідає цільовій задачі, але обмежує узагальнення результатів на весь BioASQ та інші предметні області. Зазначені 280 прогонів є перевіркою семи конфігурацій на тих самих 40 запитах, а не 280 незалежними спостереженнями.

Потребує деталізації формування пошукового корпусу та еталонів релевантності. Зокрема, необхідно пояснити, як оцінюється релевантність узагальнених вузлів дерева порівняно з первинними документами та як співвідносяться MAP@10 і використана глибина вибірки контексту Top-K = 5. Для окремого запиту у функції корисності доцільно використовувати позначення AP@10, розмежувавши його з усередненою метрикою MAP@10.

Метрики не повністю підтверджують заявлені висновки про фактологічну надійність. EC, AC та RF у наведених формулах оцінюють покриття еталонних елементів у відібраному контексті, а не їх правильне відтворення у фінальній відповіді. UER вимірює частку сутностей, відсутніх у контексті, але не виявляє всі хибні твердження про наявні сутності, зокрема неправильні дозування, заперечення або зміну напрямку взаємодії. Тому UER = 0 не є достатньою умовою цілковитої фактологічної обґрунтованості.

Статистичний аналіз є перевагою роботи, проте потребує уточнення. Нульову гіпотезу критерію Вілкоксона некоректно зведено лише до рівності математичного сподівання різниць нулю. Не описано корекції за множинні порівняння. Поліпшення Attribute Coverage та Answer F1 для окремих порівнянь статистично не підтверджені, що необхідно послідовно враховувати у висновках.

№ з/п	Характеристики та критерії оцінки рукопису наукової роботи	Рейтингова оцінка. Максимальна кількість балів (за 120-бальною шкалою)	Бали
10	Твердження про перенесення числових і реляційних зв'язків «без втрат» перевищує наведені докази. Інструкція сумаризатору зберігати факти сама по собі не забезпечує формальної гарантії їх збереження. Висока векторна подібність відповідей також недостатня для встановлення їх фактологічної еквівалентності. Доцільна незалежна експертна перевірка вилучених атрибутів, відношень і відповідей. У математичному та технічному описі наявні неузгодженості. Вектор ознак названо семивимірним, хоча одна зі складових подана чотиривимірним вектором типу запитання. На с. 13 час маршрутизації зазначено в мікросекундах, тоді як у таблиці 3.5 — 14,2 мілісекунди. Для відтворення часових результатів необхідно вказати точний ідентифікатор моделі, апаратне забезпечення, режим квантування, фактичну довжину відповідей та умови вимірювання. Практичну масштабованість не можна встановити лише за невеликим часом пошуку на дослідженій вибірці. Не наведено достатніх вимірювань зі зростанням корпусу, витрат побудови дерева та індексів. Висновки про придатність до клінічних систем потребують ширшої перевірки; результати наразі характеризують експериментальний метод пошуку й генерації. Заявлено наявність програмного стенда та репозиторію, доступного комісії за запитом. Вилучення посилання для збереження анонімності є зрозумілим, однак без супровідного коду й первинних результатів незалежна перевірка реалізації залишається невиконаною. Відомостей про власні наукові публікації або інші матеріали апробації в рукописі не наведено. Проблема сформульована чітко та обґрунтовано. Зниження балу пов'язане з тим, що окремі ідеї адаптивної гранулярності та ієрархічного RAG уже активно розвиваються в літературі. Варто детальніше описати формування набору епізодів, розподіл класів інцидентів, критерії розмітки «еталонних» рішень та процедуру запобігання упередженості при оцінюванні відповідей моделей. Варто додати порівняння з більшою кількістю сучасних RAG-рішень, зокрема з адаптивними та графовими RAG-підходами. Для клінічного впровадження потрібні додаткова валідація на інших медичних корпусах, оцінка поведінки в умовах неповних або суперечливих даних і аналіз безпеки для клінічних сценаріїв. Зниження балу пов'язане з використанням готових компонентів, що є нормою для сучасних досліджень, але потребує чіткішого розмежування внеску. Доцільно покращити наочність подання основних результатів: додати зведені графіки, таблиці порівняння конфігурацій і приклади збереження або втрати відповідних фактів. У роботі мало уваги приділено опису програмного забезпечення. Отримано конкретні кількісні результати: зниження UER, підвищення MAP@10 та обмеження накладних витрат. Для повнішої практичної валідації варто розширити оцінювання на додаткові біомедичні набори даних і реальні клінічні сценарії. Запропоновано цікавий симбіоз графа інваріантів E-A-R та адаптивного маршрутизатора гранулярності на LightGBM. Проте «збереження інваріантів» у процесі сумаризації зводиться до промпт-інжинірингу. Розмір тестової вибірки лише 40 запитів, що обмежує статистичну потужність висновків щодо роботи класифікатора. Відсутні підтвердження практичного використання. Немає публікацій. Наведено лише згадку про анонімізований репозиторій вихідного коду на GitHub. Зауваження: механізм збереження чисел і зв'язків у вузлах дерева спирається на дисципліну мовної моделі, яка отримує відповідну вказівку, а не на алгоритмічну перевірку; варто передбачити контроль того, чи справді факти потрапили до узагальнення; обсяг тестового набору замалий для впевнених узагальнень, тому формулювання висновків мають бути обережнішими, особливо щодо показників, для яких значущість не досягнута; основні метрики описують повноту знайденого контексту; фактична правильність згенерованих відповідей лікарями чи біологами не оцінювалася; є дрібні суперечності в технічному описі (параметри моделі, часові характеристики), які слід усунути для відтворюваності; публікацій немає, програмний стенд для перевірки рецензентам не надано.		
10.1	Наявність розробленого програмного забезпечення, у т.ч. веб-сервісу, мобільного застосунку	10	7,0
10.2	Рівень отриманих практичних результатів наукової роботи	10	6,8
Сума балів			88,8

Загальний висновок рекомендується для захисту на ПНПК

(рекомендується, не рекомендується для захисту на науково-практичній конференції)

Рецензент *члени Конкурсної комісії напряму «Обробка природної мови»*

«06» листопада 2026 року