Р Е Ц Е Н З І Я

Міжнародний конкурс студентських робіт зі штучного інтелекту 2026

ІІ тур
Перший етап

на наукову роботу

назва Інтелектуальна інформаційна технологія надання консультацій відвідувачам інтернет-магазину

шифр Ру-Buddy

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	9,8
2	Новизна та оригінальність ідей	15	11,0
3	Використані методи дослідження	15	11,8
4	Теоретичні наукові результати	10	6,8
5	Практична направленість результатів	20	16,8
6	Рівень використання наукової літератури та інших джерел інформації	5	4,8
7	Ступінь самостійності роботи	10	8,3
8	Якість оформлення	5	4,3
9	Наукові публікації	10	10,0
10	Недоліки роботи (пояснення зниження максимальних балів у пунктах 1-9):		

Робота виконана на високому рівні, автор демонструє глибоке розуміння проблеми достовірності відповідей консультаційних агентів і зрілий підхід до експерименту, включно з відвертим аналізом меж застосування.

Наукова новизна полягає переважно в поєднанні та адаптації відомих підходів до конкретної задачі. Сам автор обґрунтовано визнає відомість архітектури з мовною моделлю та детермінованим ядром, релаксації запитів і контрфактичного аудиту. Водночас пріоритетність запропонованої системи сертифікатів потребує точнішого порівняння з найближчими аналогами.

Основний порівняльний експеримент охоплює лише 49 тестових випадків. Базова система BM25-RAG в автономному режимі не є повноцінним LLM-консультантом, а словниковий метод індукції обмежень використовує англомовні підрядки на неангломовних каталогах. Такі порівняння демонструють окремі переваги розробки, але недостатні для висновку про її перевагу над сильними альтернативними рішеннями. Бажані абляційні експерименти та ширші незалежні набори запитів.

Теорема існування спирається на можливість зняття всіх обмежень: для непорожнього каталогу це забезпечує непорожню множину кандидатів, але саме по собі не гарантує змістовної корисності рекомендації для покупця. Потребує уточнення зв'язок цього результату із захищеними вимогами, які не можна послаблювати, та з визнаною автором вакуумністю повного зняття вимог. Для теореми 2 необхідно явно визначити функцію Regret та умови її строгої додатності для кожної нетривіальної релаксації.

Заявлені гарантії достовірності забезпечуються для детермінованого ядра та перевірюваних структурованих полів. За наведеними результатами, у вільному тексті LLM-шару частка галюцинацій становить 24,49%. Тому твердження на с. 10 і 18 про неможливість додавання моделлю непідтверджених фактів потребують узгодження з результатами розділу 3. Гарантії окремих компонентів не слід поширювати на всю відповідь консультанта.

№ з/п	Характеристики та критерії оцінки рукопису наукової роботи	Рейтингова оцінка. Максимальна кількість балів (за 120-бальною шкалою)	Бали
10	Сертифікат інваріантності довіри підтверджує стійкість до визначених комерційних втручань на дослідженій вибірці. Значення FID = 0 не є достатньою підставою для загального висновку про неупередженість системи щодо всіх комерційних чинників. Опис відтворення містить конкретні команди та назви звітів, що є перевагою роботи. Однак у наданому рукописі не наведено доступної адреси репозиторію та ідентифікатора версії програмного забезпечення. Без супровідного пакета незалежно перевірити заявлені результати й виконання 660 тестів неможливо. Представлено вагомі ознаки самостійного інженерного та дослідницького внеску: деталізований код, історію випробувань, аналіз виявлених дефектів і меж застосування. Водночас обсяг особистого внеску в опубліковані матеріали та програмну розробку доцільно уточнити під час захисту. Роботу добре структуровано, проте наявні редакційні та термінологічні похибки, зокрема повторюване словосполучення «точністю виконання», неуніфіковані іншомовні терміни та перевантаження скороченнями. Публікаційна апробація та реєстрація авторського права є позитивними характеристиками роботи. Однак для максимальної оцінки публікацій необхідно встановити, які саме результати конкурсного дослідження опубліковано. Свідоцтво про реєстрацію авторського права враховується як підтвердження оформлення програмного результату, проте не замінює оцінювання його наукової новизни й ефективності. Актуальність задачі безумовна, проте постановка має більш інженерний характер. Судячи з роботи маємо стандартний пошук найближчого сусіда за порогом, що обгорнуто теоремами, назвами та абревіатурами. Нерепрезентативна тестова вибірка і деякі неточності в методології. Гарна «студентська» спроба формалізації, проте теореми просто є обгортками білятривіальних тверджень. Проблема з рівнем галюцинацій. Робота здебільшого виконана самостійна, видне використання інструментів ШІ. Проблема сформульована чітко та обґрунтовано. Сильним є поєднання машинно-перевірюваних сертифікатів довіри, невакуумної фактичної гарантії та релаксації нерозв'язних запитів. Водночас архітектура «LLM для мови + детерміноване ядро для фактів» уже є поширеним підходом. Методологія дуже сильна. Обмеження полягає в тому, що частина ключових результатів отримана на синтетичному каталозі з еталонами, похідними від тієї самої бази. Доведено теорему існування невакуумної відповіді та формалізовано решітку релаксацій, що є суттєвим теоретичним внеском. Доцільно глибше обґрунтувати межі застосування метрики NVFG і поведінку системи на складніших багатокрокових запитах. Система має високу практичну цінність як шар перевірки для наявних LLM-консультантів інтернет-магазинів. Водночас для промислового впровадження потрібно ширше тестування на реальних каталогах конкретних торгових майданчиків. Автор провів систематичний пошук джерел. Робота демонструє високий рівень самостійності. Зниження балу пов'язане з використанням готових компонентів FastAPI, SQLite, FTS5 і зовнішніх LLM-моделей. Доцільно покращити візуальну узгодженість окремих графіків і таблиць. Розроблено повноцінну систему. Зниження підтакування на 76–85% заявлено, однак не наведено вихідних і кінцевих показників, кількості перевірених відповідей та результатів окремо за моделями. Зауваження: внесок автора варто сформулювати скромніше: архітектура, у якій мовна модель відповідає лише за мову, а факти бере детермінований модуль, сьогодні вже поширена, і самостійну цінність мають насамперед запропоновані метрики довіри; синтетичний каталог і тестові кейси, на яких отримано ключові числа, підготовлені самим автором; хотілося б побачити перевірку на зовнішньому, незалежно сформованому наборі запитів і порівняння з повноцінним LLM-консультантом; формальні теореми мають переважно допоміжний характер і за обсягом викладу дещо переважають свою змістовну вагу; для кінцевого користувача відповідь формує саме мовна модель, а її текст гейт заземлення не контролює, отже обіцяна достовірність поки що досягається лише частково; текст важко читати через велику кількість англомовних абревіатур і кальок.		
10.1	Наявність розробленого програмного забезпечення, у т.ч. веб-сервісу, мобільного застосунку	10	9,0
10.2	Рівень отриманих практичних результатів наукової роботи	10	6,0
Сума балів			98,3

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