Міжнародний конкурс студентських робіт зі штучного інтелекту 2026

II тур
Перший етап

Р Е Ц Е Н З І Я

на наукову роботу

назва Модель автономного вибору дій відновлення мікросервісної системи з урахуванням стану інфраструктури на основі машинного навчання

шифр AIRS

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	8,6
2	Новизна та оригінальність ідей	15	11,4
3	Використані методи дослідження	15	10,8
4	Теоретичні наукові результати	10	7,2
5	Практична направленість результатів	20	15,6
6	Рівень використання наукової літератури та інших джерел інформації	5	3,8
7	Ступінь самостійності роботи	10	7,0
8	Якість оформлення	5	4,0
9	Наукові публікації	10	5,2
10	Недоліки роботи (пояснення зниження максимальних балів у пунктах 1-9): Не вистачає повного опису результатів експерименту. Можливий витік інформації через action_elapsed_ms; немає оцінки якості політики вибору дій; не узгоджено 5 і 7 типів відмов та 1050 і 968 записів; неповна відтворюваність. Необхідно уточнити доступність ознак до виконання дії та підтвердити якість вибору відновлювальних дій.		
10.1	Наявність розробленого програмного забезпечення, у т.ч. веб-сервісу, мобільного застосунку	10	8,4
10.2	Рівень отриманих практичних результатів наукової роботи	10	7,2
Сума балів			89,2

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

Рецензент *члени Конкурсної комісії напряму «Обчислювальний інтелект»*

«06» листопада 2026 року

Контекстний вибір дій відновлення мікросервісів є практично значущою задачею. Поєднання телеметрії, контрольованих відмов і оцінювання альтернатив має зрозумілу мотивацію для AIOps та автономного керування інфраструктурою.

Позитивно оцінено формування емпіричного набору результатів альтернативних дій, критерій стабільного відновлення та групування записів за інцидентом. Водночас прогнозування P(успіх | контекст, дія) стандартними класифікаторами з подальшим argmax є відомою загальною конструкцією. Немає прямого зіставлення з конкретними спорідненими системами й експериментального виділення внеску запропонованих елементів.

Сильні сторони – три моделі, GroupKFold і розміщення попередньої обробки всередині Pipeline у лістингу. Проте ознака action_elapsed_ms є в NUMERIC_CANDIDATES і на графіку важливості. Якщо це фактичний час виконаної кандидатної дії, він недоступний до її вибору: виникає часовий витік інформації, якого GroupKFold не усуває. Це підтверджений склад ознак, але семантику їх отримання слід уточнити; саме завищення метрик без повторного експерименту не встановлене. Відсутні кількість фолдів, повні гіперпараметри, інтервали невизначеності та протокол вибору моделі.

Формалізовано здоровий стан, успішність відновлення, час відновлення та правило вибору дії. Однак основні моделі стандартні; нового алгоритму навчання або теоретичних гарантій не наведено. Вимогу стабільності кількох послідовних вимірювань викладено словами, але не повністю формалізовано: бракує кількості вимірів, інтервалу, тайм-ауту й конкретних порогів. Висока якість бінарного прогнозу не гарантує правильного ранжування дій для одного інциденту.

Описано контейнеризований стенд, HTTP-навантаження, моніторинг через Docker Engine і п’ять кандидатних дій, включно з NoAction. Це переконлива прикладна основа. Водночас не продемонстровано повний автономний цикл на нових інцидентах, безпечні обмеження дій і виграш у часі відновлення або доступності. Не визначено отримання fault_type та severity у реальній експлуатації, де вони не відомі з параметрів ін’єкції.

Огляд охоплює мікросервіси, ін’єкцію відмов та інтелектуальне відновлення, є нові за роком публікації.

Наявні власний опис стенду, експериментальні результати та фрагмент коду підтримують позитивну оцінку. Особистий внесок у розроблення компонентів, отримання даних і аналіз не розмежовано; повний код та історія розробки відсутні. Зниження оцінки відображає обмеженість підтвердження самостійності за рукописом, а не встановлений факт запозичення. Оцінка підлягає уточненню на захисті.

Є структура, рисунки, таблиці та додаток. Водночас у моделі відмов перелічено п’ять типів, а план і висновки використовують сім. У табл. 2 дві дії мають позначку «Так», хоча правило argmax передбачає одну. На reliability diagram присутній LightGBM, якого немає в основному порівнянні трьох моделей. Є незавершені фрази та непослідовна нумерація формул.

Наведено результати на 968 записах: для Random Forest Accuracy 92,36 %, F1 87,63 %, ROC-AUC 0,981. Матриця містить TN=632, FP=33, FN=41, TP=262; ці числа узгоджуються з Accuracy, Precision, Recall, F1 і Balanced Accuracy в табл. 3. Це позитивна внутрішня перевірка, але не незалежна валідація. Через action_elapsed_ms результат потребує повторного оцінювання для сценарію вибору дії до її виконання. Не наведено частку успішно відновлених інцидентів за обраною політикою, порівняння з правилами, найкращою сталою дією чи випадковим вибором, а також regret або виграш у RT.

Питання та рекомендації для захисту

1. Визначити момент доступності action_elapsed_ms. Якщо це вимір після запуску дії, виключити ознаку й повторити весь груповий цикл оцінювання. Якщо використовується прогнозована до дії тривалість, описати її незалежне отримання без даних відкладених інцидентів.
2. Надати паспорт вибірки: усі сім типів відмов, кількість інцидентних груп, профілі навантаження та їх розподіл, причини 82 відсутніх записів. Пояснити, чи fault × severity × replicate однозначно задає групу за різних workload_profile.
3. Пояснити відтворення початкового контексту для п’яти альтернативних дій: скидання стенду, порядок запусків і контроль впливу попереднього експерименту. Інакше результати дій можуть стосуватися різних станів.
4. На групах, виключених із навчання, обирати одну дію за argmax і вимірювати її фактичний успіх, RT та вартість. Порівняти з базовими політиками. Високий ROC-AUC прогнозу результатів окремих дій не замінює цього експерименту.
5. Уточнити табл. 2 з двома обраними діями; узгодити список моделей на рис. 8. Надати Brier score усіх моделей, спосіб побудови калібрування та підтвердження використання прогнозів для відкладених груп.
6. Надати повний код, конфігурацію стенду, дані або доступний приклад, параметри моделей, кількість фолдів, пороги успішності.

Робота має актуальну постановку, корисний експериментальний стенд і позитивні ознаки практичної реалізації. Заявлена ефективність автономного вибору дій поки підтверджена неповністю: ключовими є перевірка часової доступності ознак і окреме оцінювання політики.