DOI: 10.34185/1991-7848.itmm.2026.01.073

СТВОРЕННЯ КРИЗОВО-ЗАЛЕЖНОГО ДАТАСЕТУ ДЛЯ ADAPTIVE IRM

Березюк М.О.[1] [ORCID], Гуда А.Ю.[2] [ORCID]

[1]Український державний університет науки і технологій, аспірант, Україна

[2]Український державний університет науки і технологій,

д.т.н., професор, Україна

Анотація. *У кризових ситуаціях великі мовні моделі (LLM) мають потенціал допомагати у формуванні порад та рекомендацій, однак їх стандартна поведінка часто ігнорує специфіку події. Це знижує релевантність і може становити ризик у критичних ситуаціях. У роботі представлено підхід до створення спеціалізованого датасету для навчання та оцінки Adaptive IRM – модуля, який інжектує прихований кризовий контекст у LLM. За основу взято корпус HumAID із твітами про стихійні лиха, для яких згенеровано абстрактні запитання без прямої згадки події. Сформований набір (~41 тис. прикладів) дозволяє перевіряти, чи здатні моделі з Adaptive IRM давати відповіді, що відрізняються залежно від типу кризи, підвищуючи їх релевантність і безпечність.*

Ключові слова: *контекстно залежна генерація; великі мовні моделі; мультимодальний датасет; анотація кризових даних; кризова інформатика; датасет HumAID; ін'єкція контексту; адаптація поведінки моделі.*

Вступ

Генеративні LLM усе частіше застосовуються у кризових ситуаціях, однак без урахування контексту подій вони схильні продукувати загальні або ризиковані поради [1,2]. Відсутність у публічних корпусах пар типу «абстрактний запит – відповідь, адаптована під конкретну кризу» ускладнює оцінку методів керованої адаптації.

Запропонований кризово-залежний датасет вирішує цю проблему: він дозволяє навчати та перевіряти Adaptive IRM – модуль, що інжектує латентний контекст (наприклад, повінь, пожежа, землетрус, ураган) у LLM [3]. Завдяки цьому однакові абстрактні запитання породжують різні відповіді, сфокусовані на конкретній ситуації.

Таким чином, датасет створює основу для об'єктивної оцінки та розвитку методів контекстно залежної генерації, підвищуючи релевантність і безпечність використання LLM у критичних умовах.

Основний матеріал

Для побудови корпусу було використано відкритий датасет HumAID, який містить понад 77 тисяч твітів, промаркованих за 19 типами кризових подій, зібраних у 2016–2019 роках [4]. Це один із найбільш повних публічних ресурсів у сфері crisis informatics, що охоплює землетруси, повені, урагани, пожежі та інші стихійні лиха. З нього було відібрано приблизно 41 тисячу прикладів, що дозволило зберегти баланс між обсягом даних та дотриманням політики Twitter щодо обмеження поширення первинних записів.

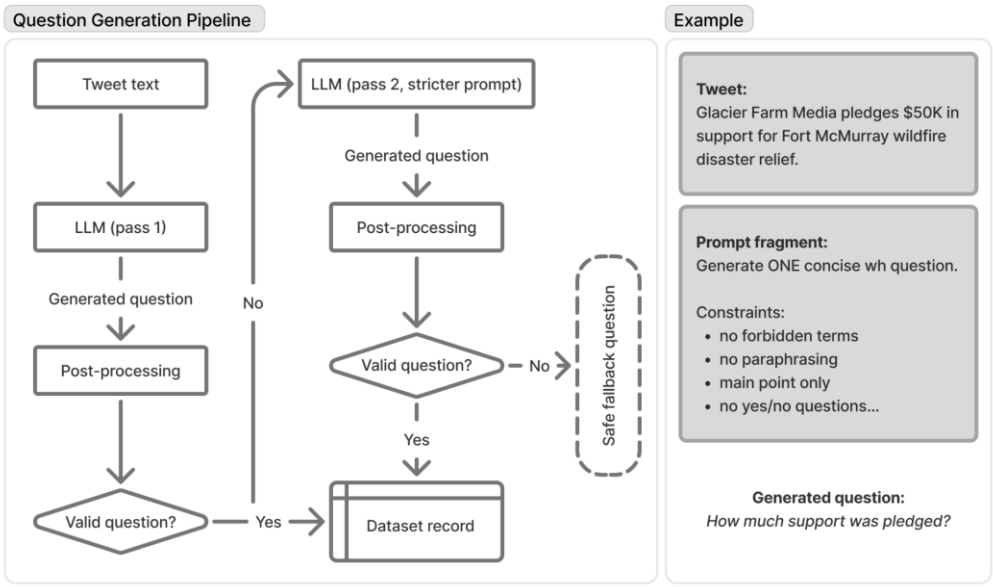

Рисунок 1 - Конвеєр генерації запитань для побудови кризо-незалежних запитів на основі твітів.

Для заданого твіту LLM генерує кандидатне запитання (перший прохід), яке далі проходить постобробку та перевірку. Якщо перевірка не пройдена, виконується другий прохід генерації з використанням суворішого промпта. Якщо обидві спроби є неуспішними, використовується безпечне запасне запитання. Запитання вважається коректним, якщо воно не містить заборонених термінів, пов’язаних із кризами, не є перефразуванням вхідного твіту, відповідає формату WH-запитання та задовольняє обмеження довжини. Постобробка включає нормалізацію, зокрема виділення одного запитання та забезпечення коректного форматування. Фрагмент промпта наведено у скороченому вигляді з ілюстративною метою.

На основі цих повідомлень за допомогою моделі GPT-OSS 20B було згенеровано абстрактні запитання, які не містять явних згадок про кризу. Для кожного твіту передбачалося одне запитання, що дозволило зменшити обчислювальні витрати та стандартизувати структуру даних. У випадках, коли модель некоректно включала назву події, використовувались підготовлені fallback-запитання на кшталт «Яка зараз ситуація?». Такий підхід дозволив зберегти узгодженість формату, а також забезпечив контроль якості: вибіркова перевірка сотні прикладів підтвердила валідність сформованих запитань.

У підсумковому датасеті кожен запис включає ідентифікатор повідомлення, його текст, категорію кризи, сформоване запитання та позначку методу генерації. Для подальших експериментів корпус поділяється на навчальну, валідаційну та тестову підвибірки зі збереженням пропорційності кризових типів, що мінімізує ризик дисбалансу та підвищує надійність оцінки.

Сконструйований датасет використовується для навчання та оцінки роботи Adaptive IRM - невеликого нейронного модуля, що інтегрується у forward pass LLM і модифікує його внутрішні представлення залежно від кризового контексту. Основна ідея полягає в тому, що на однакові абстрактні запитання модель повинна відповідати по-різному, якщо контекстом виступає, наприклад, повінь чи пожежа. Таким чином, IRM виконує роль механізму керованої адаптації, що інжектує приховану інформацію про тип події, не змінюючи основні ваги LLM.

Ефективність такого підходу оцінюється шляхом порівняння базової моделі та моделі з Adaptive IRM. Для вимірювання якості відповідей використовується метрика BERTScore, яка дозволяє обчислити семантичну близькість між згенерованим текстом та оригінальними кризовими повідомленнями. Додатково застосовується семантичний класифікатор, що дає змогу визначити, наскільки відповідь належить до правильного кризового класу. Це дозволяє оцінити не лише змістовність, а й стійкість адаптації у різних сценаріях.

Висновки

Побудований корпус із приблизно 41 тисячі пар «текст твіту – абстрактне запитання - кризовий тип» створює основу для валідації Adaptive IRM як інструменту контекстно залежної генерації. Навіть у текстовому форматі такий ресурс дозволяє перевірити, чи здатна модель змінювати свою поведінку динамічно залежно від поданого контексту.

Подальший розвиток дослідження передбачає кілька напрямів: розширення корпусу за рахунок генерації кількох запитань на один твіт, включення мультимодальних даних (зображень, сенсорних показників та інших аналітичних сигналів), а також проведення user studies із залученням експертів у сфері реагування на надзвичайні ситуації для оцінки практичної корисності та безпечності системи.

Серед ключових обмежень варто відзначити використання fallback-питань приблизно у 4.5% прикладів, а також ризик надмірної категоричності або помилкових рекомендацій з боку моделі. Для зниження цих ризиків у майбутньому планується впровадження додаткових механізмів фільтрації та аудит безпеки.

ЛІТЕРАТУРА / REFERENCE

1. Otal H. T., Canbaz M. A. C. LLM-Assisted Crisis Management: Building Advanced LLM Platforms for Effective Emergency Response and Public Collaboration // 2024 IEEE Conference on Artificial Intelligence (CAI). 2024. P. 851–859. DOI: 10.48550/arXiv.2402.10908.

2. Fengyi X. та ін. Large language model applications in disaster management: An interdisciplinary review // International Journal of Disaster Risk Reduction. 2025. Vol. 127. Art. No. 105642. DOI: 10.1016/j.ijdrr.2025.105642.

3. Березюк М. О., Гуда А. І. Контекстно залежна адаптація відповідей генеративних LLM. Інформаційні технології в металургії та машинобудуванні – ITMM'2025 : тези доп. Міжнародної наук.-техн. конф. (м. Дніпро, 23-24 березня 2025 р.). Дніпро, 2025. С. 503–508. DOI: 10.34185/1991-7848.itmm.2025.01.089.

4. Alam F., Qazi U., Imran M., Ofli F. HumAID: Human-Annotated Disaster Incidents Data from Twitter with Deep Learning Benchmarks [Електронний ресурс]. 2021. URL: https://arxiv.org/abs/2104.03090 (дата звернення: 11.03.2026).

CREATING A CRISIS-DEPENDENT DATASET FOR ADAPTIVE IRM

M. Berezuk, A. Guda

Abstract. *In crisis communications, Large Language Models (LLMs) have the potential to assist in generating guidance and recommendations; however, their default behavior often ignores the specific nature of the event. This reduces relevance and may pose risks in critical situations. This paper presents an approach to constructing a specialized dataset for training and evaluating Adaptive IRM - a module that injects latent crisis context into the forward pass of an LLM. The HumAID corpus of disaster-related tweets was used as a foundation, with abstract questions generated without explicit mentions of the crisis type. The resulting dataset (~41K examples) enables the assessment of whether models equipped with Adaptive IRM can produce responses that vary according to the crisis type, thereby improving both relevance and safety.*

Keywords: *context-aware generation; large language models; multimodal dataset; crisis data annotation; crisis informatics; HumAID dataset; context injection; model behavior adaptation.*