

ІНТЕЛЕКТУАЛЬНИЙ ПОМІЧНИК ДЛЯ АВТОМАТИЗАЦІЇ ПЕРЕВІРКИ ПРАВИЛЬНОСТІ ОФОРМЛЕННЯ АКАДЕМІЧНИХ ДОКУМЕНТІВ

Оформлення робіт студентів є невід’ємною частиною навчального процесу у закладах вищої освіти. Правильне дотримання стандартів оформлення курсових, випускних кваліфікаційних робіт має не лише формальний, але й практичний сенс, адже сприяє структурованості та зрозумілості представленої інформації. У той же час, значна кількість студентів стикається з труднощами під час оформлення своїх робіт через складність стандартів та вимог, що часто змінюються.

Сучасні технології штучного інтелекту, зокрема великі мовні моделі (LLM), відкривають нові можливості для автоматизації цього процесу. Використання LLM, таких як ChatGPT, дозволяє створювати кастомні рішення для перевірки текстових документів на відповідність встановленим стандартам. Використання сучасних технологій обробки природної мови (NLP) дозволить створити систему, яка автоматично аналізуватиме структуру, форматування та бібліографічні елементи статей, забезпечуючи високу ефективність і відповідність стандартам. Така система може суттєво зменшити навантаження на викладачів, забезпечуючи швидку та точну перевірку оформлення ВКР, а також надати студентам зрозумілі рекомендації щодо виправлення помилок.

Метою цієї статті є опис розробки кастомного GPT – інтелектуального контролера для автоматизації перевірки правильності оформлення курсових та випускних кваліфікаційних робіт відповідно до українських державних стандартів. У статті будуть розглянуті технологічні основи системи, її структура, функціонал та результати тестування.

Традиційний процес перевірки правильності оформлення студентських робіт є трудомістким, часомістким, рутинним і залежить від людського фактору. Автоматизація цього процесу дозволить значно підвищити продуктивність самого процесу перевірки та підвищити якість оформлення робіт студентів.

Великі мовні моделі стали революційним інструментом у сфері обробки природної мови, відкриваючи нові можливості для автоматизації завдань, пов'язаних з текстом. Для реалізації інтелектуального помічника для перевірки правильності оформлення робіт студентів доцільно використовувати вже існуючі готові до використання моделі типу GPT. На сьогоднішній день можна вибудовувати систему перевірки, спираючись на локальні або доступні on-line моделі [1; 2].

Дослідження присвячені використанню LLM для задач обробки документів показали значний їх потенціал у покращенні точності і релевантності опрацювання текстів [3; 4]. Використання локальних і онлайн LLM для створення інтелектуального помічника має свої переваги та недоліки, які слід враховувати залежно від специфіки завдань та вимог користувача.

Локальні LLM мають кілька вагомих переваг. По-перше, вони забезпечують високий рівень конфіденційності даних, оскільки вся інформація обробляється локально, без необхідності передавати її через Інтернет. По-друге, локальні моделі працюють автономно, що робить їх незалежними від наявності стабільного Інтернет-з'єднання. Також вони дозволяють користувачам мати повний контроль над налаштуваннями та оптимізацією моделі, що дає змогу адаптувати її під специфічні потреби організації. Однак локальні LLM мають і певні недоліки. Основним є високі вимоги до обчислювальних ресурсів, необхідних для розгортання та підтримки таких моделей, що може бути надто обтяжливим для невеликих організацій. Крім того, підтримка та оновлення локальних моделей потребує значних технічних знань і ресурсів. Ще одним недоліком є те, що локальні моделі можуть застарівати без регулярних оновлень, що може впливати на їхню ефективність і точність.

Онлайн LLM, у свою чергу, мають свої переваги. Вони зручні у використанні, оскільки доступні з будь-якої точки світу через Інтернет і не потребують значних локальних ресурсів. Онлайн моделі регулярно оновлюються розробниками, що забезпечує актуальність та високу якість результатів. Такі моделі легко інтегруються у робочі процеси, що робить їх ідеальними для організацій, які потребують швидкого доступу до сучасних технологій без великих початкових інвестицій, до яких однозначно можна віднести державні ВНЗ. Втім, онлайн LLM мають і недоліки. Основним є залежність від стабільного Інтернет-з'єднання, без якого робота таких моделей неможлива. Крім того, передача даних через Інтернет створює ризики витоку конфіденційної інформації, що потребує ретельного контролю за безпекою. Онлайн моделі також менш гнучкі

в налаштуванні, оскільки контроль над моделлю залишається у постачальника, що може обмежувати можливості адаптації до специфічних потреб користувача.

Для реалізації пілотного проєкту було вирішено побудувати інтелектуального помічника, що відслідковує коректність заповнення титульної сторінки курсової роботи. В якості технологічного ядра системи вирішено використати можливості, які надає ChatGPT від OpenAI для побудови кастомних моделей.

На підготовчому етапі для створення бази правил, на яку буде спиратись помічник, були

- завантажені стандарти з оформлення технічної документації;
- визначені правила і шаблон для перевірки текстових елементів титульної сторінки;
- завантажені приклади вірно заповнених титульних сторінок.

Цікаво, що в ході роботи помічник самостійно зробив логічну структурування вимог і сформулював відповідні групи (концепти): «Відповідність шаблону», «Формат змінних даних», «Форматування тексту», «Підписи» та інші. Крім того, модель сформувала і додала власні правила. Так, жодних обмежень до року виконання роботи не висувалось, але після певної кількості ітерацій навчання при завантаженні титульної сторінки з роботи попередніх років, помічник самостійно згенерував зауваження «Рік виконання: У полі вказано рік «2021», але цей рік не відповідає актуальній даті. За документами курсові роботи, як правило, оформлюються у поточному році.».

**Повний список правил для перевірки оформлення титульної сторінки курсової роботи:**

- Відповідність тексту шаблону:**
 - Постійний текст у шаблоні має залишатися незмінним.
 - Змінні дані (текст у прямокутних дужках у шаблоні) повинні бути заповнені індивідуальними даними.
- Формат змінних даних:**
 - Змінні дані повинні бути виконані у тому самому форматі (тип шрифту, кегль, міжрядковий відступ), що й текст у прямокутних дужках у шаблоні.
- Форматування тексту:**
 - Шрифт має бути Times New Roman.
 - Розмір шрифту — 14 пунктів.

Рисунок 1 – Угрупування правил, виконане мовною моделлю

Таким чином, можна стверджувати, що використання сучасних технологій штучного інтелекту, зокрема кастомних помічників побудованих на великих мовних моделях, для автоматизації процесу перевірки академічної документації забезпечить значне зростання ефективності та якості даного процесу.

У роботі запропоновано підхід, який поєднує потенціал передових методів великих мовних моделей зі спеціалізованими правилами оформлення курсових та випускних кваліфікаційних робіт студентів.

Пілотний проєкт продемонстрував ефективність використання GPT для автоматизації перевірки ВКР, де модель не лише перевіряла відповідність стандартам, а й самостійно формувала додаткові правила, що підвищувало якість і точність перевірки. Це підкреслює потенціал таких технологій у покращенні академічних процесів.

Подальші дослідження можуть бути спрямовані на розробку та впровадження більш просунутих кастомних моделей з удосконаленими алгоритмами обробки контексту, що дозволить розширити функціонал помічника для перевірки структури (зміст, вступ, висновки) та відповідності стилю та формату (цитування, списки літератури), інш.

Література:

1. Sufi, Fahim GPT in Research: A Systematic Review on Data Augmentation. Information. 2024. 15. 99. 10.3390/info15020099
2. Baviskar, Dipali & Ahirrao, Swati & Potdar, Vidyasagar & Kotecha, Ketan. Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. 2021. IEEE Access. Pp. 1-1. 10.1109/ACCESS.2021.3072900
3. Миколайчук Роман & Миколайчук Аліса. Використання технологій штучного інтелекту для автоматизації процесу обробки документів. Сучасні інформаційні технології у сфері безпеки та оборони. 2023. № 50. Р. 111–117. 10.33099/2311-7249/2024-50-2-111-117
4. Mahadevkar, S. V., Patil, S., Kotecha, K. et al. Exploring AI-driven approaches for unstructured document analysis and future horizons. J Big Data. 2024. 11, 92. <https://doi.org/10.1186/s40537-024-00948-z>