



ЗАТВЕРДЖУЮ

Проректор з наукової роботи
Національного університету
«Львівська політехніка»

Іван ДЕМИДОВ

25 " жовтня 2024 р.

Висновок

про наукову новизну, теоретичне та практичне значення результатів
дисертації «Методи та засоби багатоетапної генерації семантично коректних текстів»
здобувача наукового ступеня доктора філософії за спеціальністю
124 - Системний аналіз (галузь знань 12 - Інформаційні технології)
Здебського Петра Васильовича
наукового семінару кафедри інформаційних систем та мереж

1. Актуальність теми дисертації

Завдання узгодження великих мовних моделей або проблема контролю штучного інтелекту є особливо актуальною сьогодні у зв'язку з тим, що вони стають більш ефективними у вирішенні різного роду завдань. Здатність розв'язування складних завдань призводить до того, що моделі можуть "обійти" задачу. Тобто, формально вони розв'язують поставлене перед ними завдання, але не так, як від них очікують дослідники. Прикладом у контексті робототехніки може бути спроба робота ввести в оману оцінювача, розташувавши маніпулятор між камерою та об'єктом, щоб вдати, що він його бере.

Проблема узгодженості виникає через обмеженість дослідників у формулюванні тих складних завдань, які сучасні мовні моделі можуть вирішувати. Упустивши часткові випадки використання при постановці задачі може призвести до того, що моделі штучного інтелекту вироблять поведінку, щоб обманювати оцінювачів, а не виконувати завдання. Також ефективність таких алгоритмів залежить від інтуїції оцінювача щодо того, яка поведінка вважається правильною.

Ця проблема є небезпечною при використанні системи у реальних умовах, де помилка є особливо чутливою для кінцевого користувача, бо важко контролювати текст, що генерує модель і немає гарантій стосовно того, що текст буде співпадати із очікуваннями користувача. Наприклад, текст, створений моделлю GPT-3, іноді може містити недостовірну, токсичну або шкідливу інформацію. В такому випадку модель можна назвати узгодженою із потребами користувача. Великі мовні моделі часто зустрічаються з явищем, що називається галюцинаціями. Це означає, що вони можуть видавати неправдиві відповіді або вигадувати факти. Це може призвести, наприклад, до генерації неправдивих новин, навіть якщо модель була натренована генерувати лише достовірні результати.

Провідна компанія у цій сфері OpenAI визначає цей напрям як один із головних. Вони використовують запитання, які вони отримали від користувачів їхнього API, щоб тренувати свої моделі, що краще узгоджуються із реальними запитами користувачів.

2. Зв'язок теми дисертації з державними програмами, науковими напрямами університету та кафедри

Тема дисертації відповідає науковому напрямку кафедри “Дослідження, розробка та впровадження Internet-застосувань, інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, проведення оцінки наявних технологій та знань з метою прискорення процесів формування сучасного інформаційного суспільства” та науковому напрямку університету “Перспективні комп'ютерні системи та інформаційні технології”.

3. Особистий внесок здобувача в отриманні наукових результатів

Основні положення та результати дисертаційної роботи одержані автором самостійно. Особисто здобувачеві належать наступні наукові результати: розроблено модель процесу багатоетапного генерування тексту, метод генерування тексту з використанням дискримінатора для перевірки узгодженості результатів із запитом користувача, проведено експериментальне дослідження точності запропонованого методу у порівнянні із звичайним використанням великої мовної моделі.

4. Достовірність та обґрунтованість отриманих результатів та запропонованих автором рішень, висновків, рекомендацій

Обґрунтовано вибір стратегії досліджень з використанням великої мовної моделі, а також конструювання промпту для реалізації запропонованого методу. Було перевірено експериментальним шляхом покращення точності запропонованого методу у порівнянні із звичайним використанням великої мовної моделі.

5. Ступінь новизни основних результатів дисертації порівняно з відомими дослідженнями аналогічного характеру

Вперше розроблено: модель генерування текстів яка виконує повторну перевірку згенерованого тексту, що дало можливість підвищити рівень семантичної коректності результату; вперше розроблено метод генерування тексту, що включає дискримінатор для перевірки узгодженості результатів із запитом користувача, що дало змогу підвищити точність генерування. Удосконалено метод валідації згенерованого тексту, що на відміну від відомих методів, не вимагає зворотного зв'язку людини, що дало змогу повністю автоматизувати процес генерування.

6. Перелік наукових праць, які відображають основні результати дисертації

Статті у наукових фахових виданнях України:

1. Здебський П. В., Берко А. Ю. Метод покращення якості генерування тексту за рахунок повторного передавання згенерованого тексту на модель. Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. № 1 (331). С. 259–263. Особистий внесок автора: розроблено метод генерування текстового контенту із формуванням набору прикладів для повторної передачі на велику мовну модель та перевірено його точність.

2. Здебський П. В., Берко А. Ю. Перевірка тексту після генерації за допомогою великих мовних моделей для фільтрування неправильних відповідей. Наукові праці Вінницького національного технічного університету. 2024. № 1. Особистий внесок автора: розроблено метод для відсіювання неправильно згенерованого тексту та перевірено його точність.

3. Берко А. Ю., Здебський П. В., Висоцька В. А. Метод формування набору даних для перевірки якості вивчення мовними моделями транзитивного відношення у контексті задачі логічного висновку. Вісник Національного університету “Львівська політехніка”. Серія: Інформаційні системи та мережі. 2024. Вип. 14. С. 46–60. Особистий внесок автора: розроблено метод формування набору даних із транзитивними відношеннями у контексті задачі логічного висновку.

Статті у виданнях інших держав:

1. Zdebskyi P., Berko A., Vysotska V. Investigation of transitivity relation in natural language inference. CEUR Workshop Proceedings. 2023. Vol. 3396 : Computational linguistics and intelligent systems 2023 : proc. of the 7th Intern. conf. on computational linguistics and intelligent systems. Vol. II: Computational linguistics workshop, Kharkiv, Ukraine, Apr. 20-21, 2023. P. 334–345. Особистий внесок автора: досліджено точність великих мовних моделей у контексті транзитивних відношень на задачі логічного висновку.

Матеріали конференцій:

1. Zdebskyi P., Berko A., Vysotska V., Voloshyn S., Chyrun S., Schuchmann V. Framework for improving the effectiveness of discussions at English-language articles analysis. Advanced information and communication technologies : proc. of the 5th IEEE Intern. conf., Lviv, Ukraine, Nov. 21–25, 2023. 2023. P. 127–132. Особистий внесок автора: розроблено систему, що включає існуючу попередньо навчену велику мовну модель для підтримки дискусій.

Основні положення та результати роботи викладено у повному обсязі у перелічених наукових працях.

7. Апробація основних результатів дослідження на конференціях, симпозіумах, семінарах тощо

Основні наукові та практичні результати роботи доповідались та обговорювались на таких міжнародних конференціях:

1. Computational linguistics and intelligent systems 2023: proc. of the 7th Intern. conf. on computational linguistics and intelligent systems. Vol. II: Computational linguistics workshop, Kharkiv, Ukraine, Apr. 20-21, 2023.

2. Advanced information and communication technologies: proc. of the 5th IEEE Intern. conf., Lviv, Ukraine, Nov. 21–25, 2023.

Також, результати дисертаційних досліджень регулярно доповідалися на наукових семінарах кафедри “Інформаційні системи та мережі” Національного університету “Львівська політехніка”.

8. Наукове значення виконаного дослідження із зазначенням можливих наукових галузей та розділів програм навчальних курсів, де можуть бути застосовані отримані результати

Результати дослідження можуть бути застосовані у галузі “Інформаційні технології”, а саме у навчальному курсі “Обробка мови методами штучного інтелекту” рівня бакалавр спеціальності “Комп’ютерні науки” спеціалізації “Системи штучного інтелекту”.

9. Практична цінність результатів дослідження із зазначенням конкретного підприємства або галузі народного господарства, де вони можуть бути застосовані

Запропонований підхід буде особливо корисним у завданнях, що пов'язані з генеруванням тексту, де існує багато варіантів правильно згенерованого тексту і його правильність легко перевірити. Наприклад, у чат-бот асистентах, що можуть використовуватись працівниками різних галузей, щоб швидко отримати потрібну їм інформацію і тим самим підвищити їхню продуктивність.

10. Оцінка структури дисертації, її мови та стилю викладення

Дисертація відповідає вимогам до оформлення дисертації згідно із наказом Міністерства освіти і науки України № 40 від 12 січня 2017 р.

У ході обговорення дисертації до неї не було висунуто жодних зауважень щодо самої суті роботи.

11. З урахуванням зазначеного, на науковому семінарі кафедри Інформаційних систем та мереж ухвалили:

11.1. Дисертація Здебського Петра Васильовича «Методи та засоби багатоетапної генерації семантично коректних текстів» є завершеною науковою працею, у якій розв'язано конкретне наукове завдання: «Розробка методів і засобів підвищення семантичної коректності текстового контенту згенерованого великими мовними моделями», що має важливе значення для інформаційних технологій.

11.2. Основні наукові положення, методичні розробки, висновки та практичні рекомендації, викладені у дисертаційній роботі, логічні, послідовні, аргументовані, достовірні, достатньо обґрунтовані. Дисертація характеризується єдністю змісту.

11.3. У 9 наукових публікаціях повністю відображені основні результати дисертації, з них 4 статті у наукових фахових виданнях України та 3 статті у наукових періодичних виданнях інших держав; 2 статті у виданнях України, які входять до міжнародних наукометричних баз.

11.4. Дисертація відповідає вимогам наказу МОН України № 40 від 12.01.2017 р. «Про затвердження вимог до оформлення дисертації», Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії (Постанова Кабінету Міністрів України від 12 січня 2022 р. № 44, зі змінами).

11.5. Дисертація є результатом самостійних досліджень, не містить елементів фальсифікації, компіляції, плагіату та запозичень, що констатує відсутність порушення академічної доброчесності. Використання текстів інших авторів мають належні посилання на відповідні джерела.

11.6. З урахуванням наукової зрілості та професійних якостей Здебського П. В. дисертація «Методи та засоби багатоетапної генерації семантично коректних текстів» рекомендується для подання до розгляду та захисту у спеціалізованій вченій раді.

За затвердження висновку проголосували:

за - 40
проти - 0
утримались - 0

Головуючий на науковому семінарі
кафедри Інформаційних систем та мереж
зав. кафедри ІСМ, д.т.н., професор

 Василь ЛИТВИН

Рецензенти:

д.т.н., ст. н. с., професор кафедри ІСМ

 Дмитро ДОСИН

к.т.н., доцент, доцент кафедри ІСМ

 Тетяна ШЕСТАКЕВИЧ

Відповідальний у ННІ за атестацію PhD
к.т.н., доцент, доцент кафедри АСУ

 Анатолій БАТЮК

"19" квітня 2024 р.