

Голові спеціалізованої вченої ради

ДФ 74.122.2025

Одеського національного університету імені І.І. Мечникова

доктору технічних наук, професору

Мещерякову Володимирі Івановичу

Відгук

офіційного опонента Арсирій Олени Олександрівни,

доктора технічних наук, професора,

завідувачки кафедри інформаційних систем

Національного університету «Одеська політехніка»

про дисертацію

Бочарової Майї Юріївни

**«Моделі та інформаційна технологія предметно-орієнтованої обробки
природномовних текстів»,**

подану на здобуття наукового ступеня доктора філософії

за спеціальністю 122 Комп'ютерні науки,

галузь знань 12 Інформаційні технології

(Одеса, 2025)

1. Актуальність дисертаційного дослідження.

Дисертаційне дослідження є актуальними та має високу наукову і прикладну цінність у контексті сучасних тенденцій цифровізації кадрових процесів. У світі, де обсяг даних резюме кандидатів і вакансій роботодавців зростає експоненційно, особливо важливим є створення інтелектуальних систем, здатних швидко й точно зіставляти ці масиви інформації. Але проведений здобувачкою аналіз предметної області показав, що більшість існуючих автоматизованих систем добору персоналу не враховують семантичні аспекти контексту резюме та вакансій, що призводить до недооцінки компетенцій потенційно цінних кандидатів або взагалі пропуску їх резюме. Запропонована технологія предметно-орієнтованої обробки природномовних текстів на основі трансформерних моделей дозволяє значно покращити якість зіставлення вакансій і резюме.

Крім того, дослідження має потенціал для подальшого масштабування, наприклад, у напрямках планування кар'єри, побудови індивідуальних траєкторій розвитку працівників, а також прогнозування потреб у навчанні. Це робить роботу не лише актуальною, але й перспективною.

2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації.

Аналіз змісту дисертаційної роботи і наукових публікацій Бочарової Майї дають змогу зробити висновок про те, що основні наукові положення, висновки і рекомендації, які сформульовано в дисертаційній роботі, відповідають темі досліджень, їх вірогідність забезпечено коректністю визначення об'єкту та предмету дослідження, які узгоджено з метою і задачами наукового дослідження.

Дисертаційна робота спрямована на вирішення складної комплексної задачі, яка знаходиться на перетині наукових сфер обробки природної мови, штучного інтелекту та управління людськими ресурсами. Враховуючи стрімке впровадження технологій штучного інтелекту у галузь рекрутингу, застосування трансформерних моделей, методів векторної дистиляції та фразового подання є обґрунтованими для обробки природномовних текстів резюме та вакансій.

Обґрунтованість отриманих теоретичних результатів дисертаційного дослідження базується на коректному використанні інструментарію системного аналізу і моделювання та підкреслюється цілісністю дослідницької логіки від удосконалення існуючих та розробки нових моделей і методів інтелектуальної обробки HR-документів до створення нової інформаційної технології з повним технічним стеком (сегментація, парсинг, сумаризація, векторизація). При цьому авторка системно аналізує вплив кожного етапу обробки на кінцеву якість рекомендацій зіставлення текстових

документів резюме та вакансій користуючись релевантних метриках продуктивності та точності.

Всі припущення при моделюванні, зазначені в роботі, є загальноприйнятими та узгодженими з розробками в сфері комп'ютерних наук, що підтверджуються всебічним аналізом значного переліку літературних джерел з досліджуваної теми.

Обґрунтованість наукових положень, висновків та рекомендацій також підтверджується даними експериментальних досліджень та отриманими практичними результатами. Суттєвою перевагою дисертації є експериментальна перевірка гіпотез дослідження з використанням даних реальних компаній, що надає висновкам роботи не тільки обґрунтованість, но беззаперечну прикладну значущість

3. Наукова новизна одержаних результатів.

Дисертація Бочарової Майї є завершеною кваліфікаційною науковою роботою, що присвячена розв'язанню актуальних науково-практичних задач зі створення моделей, методів та інформаційної технології предметно-орієнтованої обробки природномовних текстів на кшталт HR-документів ґрунтуючись на підході, який поєднує спеціалізовані механізми для роботи з нестандартними форматами візуально насичених документів, методи подання фраз у контексті, структуроване некероване навчання та сумаризацію довгих текстів документів.

У дисертаційній роботі сформульовано і обґрунтовано низку теоретичних положень, пропозицій та висновків, які містять ознаки наукової новизни, а саме:

- *удосконалено* модель подання токенів для візуального насичених документів, яка відрізняється від існуючих безпосередньою інтеграцією параметрів про стильові ознаки, що дозволяє підвищити якість подань токенів таких документів без використання методів комп'ютерного зору;

– *вперше запропоновано* метод подання фраз у контексті, що базується на використанні спеціальних маркерів для виділення фраз, які моделюються, що дозволяє значно пришвидшити процес подання фраз за рахунок використання лише однієї моделі та покращити якість в порівнянні з базовими методами;

– *вперше запропоновано* метод структурування документа для некерованого навчання подань текстів у сфері управління персоналом, що дозволяє адаптувати модель до домену та як наслідок підвищити якість сумаризації документів;

– *удосконалено* метод зменшення обсягу тексту, що оброблятиметься моделлю, який відрізняється від існуючих урахуванням структури документа та ключових фраз, для подальшої сумаризації, що дозволяє підвищити якість сумаризації довгих документів;

– *вперше запропоновано* модель векторизації документів у сфері управління людськими ресурсами на основі векторних подань секцій та механізму самоуваги разом із абсолютним позиційним кодуванням, що дозволяє покращити якість подань документів, довжина яких перевершує ліміт токенів, притаманний моделі.

Положення наукової новизни, які виносяться на захист сформульовані відповідно до тематики дослідження, відповідають меті дисертації, виконані повністю та підкреслюють, що здобувачка володіє методологією наукової діяльності повною мірою.

У тексті дисертації відсутні порушення академічної доброчесності.

4. Повнота викладу основних результатів дисертації в опублікованих публікаціях, зарахованих за темою дисертації.

Основні результати дисертаційної роботи викладено в 8 публікаціях, з них: у 1-й, статті, опублікованій у виданні, що проіндексовано у наукометричній базі Scopus; у 3-х статтях, які опубліковано у виданнях, включених до Переліку фахових видань України (всі видання індексуються у

міжнародних наукометричних базах даних, зокрема: Index Copernicus International, Ulrich's Periodicals Directory, Electronic Journals Library, Google Scholar), а також у 4-х тезах Міжнародних наукових конференцій (Малазія, Україна).

Аналіз опублікованих робіт показав, що основні положення наукової новизни, які виносяться на захист повністю викладено в зазначеному в дисертації переліку праць

5. Практичне значення одержаних результатів.

Практичну цінність дисертаційної роботи доводить широкий спектр впроваджень в ІТ-компаніях та освітньому процесі. Розроблені методи та архітектури були апробовані й інтегровані у виробничі середовища компаній Daxtra Technologies та Data Science UA, що свідчить про їх прикладну ефективність та технологічну відповідність потребам реального бізнесу. В актах впровадження наданих цими компаніями зазначено, що особливе значення має побудова технологічного процесу зіставлення резюме та вакансій на основі трансформерної архітектури із введенням модуля сумаризації. Це дозволяє автоматизувати етап попереднього аналізу, зменшити навантаження на рекрутера та підвищити точність відбору кандидатів.

Впровадження некерованих методів навчання для формування векторних подань посад, а також методів подання фраз у контексті сприяє підвищенню якості класифікації та нормалізації в HR-аналітиці. Додатково, створення еталонних датасетів та їх публікація у відкритому доступі є значним внеском у розвиток ІТ-галузі, оскільки забезпечує можливість верифікації та повторного використання розробок науковою спільнотою.

Також акти засвідчують, що результати, які отримано в дисертаційній роботі, використано при виконанні науково-дослідної роботи «Методи, моделі, інформаційні технології розподілених систем підтримки прийняття організаційних рішень» та розробці лекційної та практичної частини курсу з дисципліни: «Методи обробки текстів природної мови» на кафедрі

6. Дискусійні положення та зауваження щодо змісту дисертації.

1. В роботі на сторінці 23 представлений занадто великий перелік поставлених задач, що створює враження їх фрагментарності та розпорошеності. Крім того це ускладнює аналіз виконання цих задач для опонента. Наприклад, вираз «інформаційну теорію багатомовної дистиляції текстових подань» із задачі 9 взагалі не знайдено по тексту дисертації.
2. Розділ 2 дисертаційної роботи присвячений створенню 3-х методів обробки природномовних текстів (згідно із науковою новизною). Доцільно було б описати методи у вигляді послідовності кроків виконання. Крім того при викладені методів недостатньо повно описані обмеження їх використання щодо обчислювальної складності та використання пам'яті при збереженні векторних представлень документів.
3. Схема інформаційної технології предметно-орієнтованого аналізу природномовних текстів AI ResJobFit, яку представлена на сторінці 91 потребує пояснень щодо наявності двох однакових паралельних послідовностей кроків та відсутності блока прийняття рішень.
4. Не заважаючи на той факт, що набір пов'язаних між собою метрик для оцінки якості сумаризації ROUGE (сторінка 103) є найбільш популярним доцільно було б навести також і інші автоматизовані метрики оцінки, а також вказати на можливість експертних оцінок (Human Evaluation) та спеціалізованих завдань (QA-based Evaluation).
5. Виходячи з тексту дисертаційної роботи не зовсім зрозуміло, які саме програмні засоби та технології використано для реалізації прототипу. В тексті описано вибір параметрів, але не ясно, які саме бібліотеки та мова програмування використовуються для реалізації методів та підходів.

В цілому, вказані зауваження не впливають на високий науковий самого проведеного дисертаційного дослідження

7. Висновок про відповідність дисертації вимогам.

Дисертаційна робота Бочарової Майї Юріївни «Моделі та інформаційна технологія предметно-орієнтованої обробки природномовних текстів» є цілісним, самостійним, завершеним науковим дослідженням, яке має науково-прикладне значення, тема і зміст якого відповідає спеціальності 122 «Комп'ютерні науки».

З огляду на актуальність теми дисертації, обґрунтованості наукових положень, висновків і рекомендацій, які сформульовані в дисертації, їхньої новизни, практичної цінності, повноти викладення в наукових публікаціях, відсутності порушень академічної доброчесності, вважаю що дисертація «Моделі та інформаційна технологія предметно-орієнтованої обробки природномовних текстів» відповідає вимогам «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 року №44, із змінами, а її авторка *Бочарова Майя Юріївна* заслуговує на присудження їй наукового ступеня доктора філософії у галузі знань 12 — Інформаційні технології за спеціальністю 122 — Комп'ютерні науки.

Офіційний опонент,

завідувачка кафедри інформаційних систем

Національного університету «Одеська політехніка»

доктор технічних наук, професор,

Олена Арсірій