РЕЦЕНЗІЯ
на дисертаційну роботу
Кравчука Олександра Анатолійовича
на тему «**Метод оцінювання ризиків життєвого циклу моделей штучного інтелекту геопросторового сприйняття для інформаційних підсистем роїв безпілотних літальних апаратів»**,
представлену на здобуття ступеня доктора філософії
в галузі знань 12 «Інформаційні технології»
за спеціальністю 122 «Комп'ютерні науки»

Актуальність теми дисертації

Актуальність дисертаційного дослідження зумовлена стрімким розширенням застосування методів штучного інтелекту для керування безпілотними літальними апаратами, зокрема для вирішення задач розпізнавання об'єктів, формування карт середовища, навігації та координації дій у складі рою БпЛА. У ройовій системі на відміну від одиничного автономного агента, помилка окремої ШІ-моделі може поширюватися через обмін даними, спільні карти та механізми координації, впливаючи на колективну поведінку системи.

Останні досягнення мікроелектроніки завдяки мініатюризації дозволяють насичувати безпілотні системи сучасними датчиками, які генерують великі обсяги даних, таких як фотозображення, потокове відео та звук, але не знімають обмеження на обчислювальні, енергетичні та комунікаційні ресурси БпЛА, що набуває особливої важливості в динамічному середовищі в умовах реального часу.

Як наслідок виникають проблеми, пов'язані з управлінням ризиками штучного інтелекту, дослідженням моделей геопросторового сприйняття та безпечністю автономних багатоагентних систем. Ці питання розглядаються світовою наукою, але переважно – окремо, не пропонуючи єдиного підходу, який би одночасно структурував ризики за стадіями життєвого циклу ШІ-моделі, враховував людські та технологічні фактори, забезпечував їх кількісне пріоритизування, перевірку стійкості отриманих оцінок і, що особливо важливо для ройових систем, враховував би системні наслідки поширення локальних помилок сприйняття між агентами.

Таким чином, актуальність роботи визначається не лише практичною потребою підвищення надійності ШІ-керованих БпЛА, але й наявністю чітко окресленої наукової прогалини на стику управління ризиками штучного інтелекту, моделей геопросторового сприйняття та колективної поведінки багатоагентних систем. Розроблення методу комплексного оцінювання таких ризиків, який дозволяє перейти від їх загального переліку до структурованого кількісного пріоритизування та подальшого планування валідації й заходів

зниження ризиків, є актуальним і науково обґрунтованим завданням у галузі комп'ютерних наук.

Ступінь обґрунтованості наукових положень, висновків і рекомендацій

В дисертації розв'язано наукове завдання з розроблення методу комплексного оцінювання ризиків, що виникають на різних стадіях життєвого циклу ШІ-моделей геопросторового сприйняття для інформаційних підсистем роїв БпЛА, з урахуванням можливості поширення в ройовій системі локальних помилок сприйняття через взаємодію агентів і формування внаслідок цього системних ризиків порушення колективної поведінки рою.

Об'єктом дослідження є процес комплексного оцінювання ризиків, що виникають на різних стадіях життєвого циклу ШІ-моделей геопросторового сприйняття для інформаційних підсистем роїв БпЛА. Предметом дослідження – метод комплексного оцінювання цих ризиків з урахуванням можливих системних наслідків поширення локальних помилок сприйняття, а також моделі, критерії, показники та процедури, які забезпечують його реалізацію і перевірку результатів оцінювання.

Обґрунтованість наукових положень, висновків і рекомендацій забезпечується послідовним переходом від аналізу стандартів і наукових джерел до формування вимог до методу, побудови таксономії, застосування методу аналізу ієрархій, аналізу стійкості результатів пріоритизування, експериментального дослідження просторово-геометричних помилок і симуляції їх поширення у графі взаємодії агентів. Автор розмежовує сценарні результати пріоритизування та результати перевірки проявів ризиків. Також прямо зазначено, що змодельовані профілі суджень не є експертним консенсусом, а дослідження мультимодальних систем загального призначення, тобто систем, що одночасно оперують даними різної природи, не дає вичерпної оцінки спеціалізованих моделей геопросторового сприйняття.

Наукова новизна одержаних результатів

Наукова новизна одержаних результатів на думку рецензента полягає в наступному.

1. Вперше розроблено метод комплексного оцінювання ризиків, що виникають на різних стадіях життєвого циклу ШІ-моделей геопросторового сприйняття для інформаційних підсистем роїв БпЛА, який відрізняється поєднанням структурування та класифікації ризиків за стадіями життєвого циклу, їх кількісного пріоритизування, аналізу стійкості результатів оцінювання та врахування можливих системних наслідків локальних помилок сприйняття, що дозволяє ідентифікувати й класифікувати ризики, перевіряти

стійкість результатів оцінювання та враховувати наслідки можливих локальних помилок сприйняття.

2. Вперше розроблено таксономію ризиків ШІ-моделей геопросторового сприйняття для роїв БпЛА, яка віжрізняється пов’язуванням ризиків зі стадіями селекції знань, аналізом знань, навчанням й експлуатацєю моделі та їх розподілом за людськими і технологічними факторами, що спрощує здійснення як первинного кількісного пріоритизування, так ведення деталізованого реєстру.

3. Вдосконалено процедуру кількісного пріоритизування ризиків на основі методу аналізу ієрархій шляхом використання кількох профілів змодельованих експертних суджень і моделювання Монте-Карло з контролем узгодженості матриць попарних порівнянь, що дозволяє перевіряти стійкість результатів пріоритизування та виокремлювати ризики, які зберігають високий пріоритет за різних профілів суджень.

Експериментальне дослідження мультимодальних ШІ-систем і симуляція поширення локальної помилки не подаються в роботі як положення наукової новизни.

Практичне значення одержаних результатів

Практичне значення здобутих результатів полягає в можливості застосування розробленого методу для відтворюваного оцінювання ризиків ШІ-моделей геопросторового сприйняття, їх пріоритизування та планування подальшого тестування, валідації й моніторингу.

Конкретними прикладними результатами є ядро таксономії з 20 ризиків для первинного оцінювання та розширена таксономія з 71 ризику для ведення деталізованого реєстру. Для ризиків визначено стадії виникнення, категорії факторів, механізми прояву, індикатори раннього попередження, заходи зниження та способи перевірки. Практичну спрямованість мають також протоколи експериментальної і симуляційної перевірки, відкриті програмні матеріали та рекомендації щодо контролю невизначеності й обмеження каскадного поширення помилок.

Зазначені результати створюють основу для визначення пріоритетних напрямів тестування та зниження ризиків, але не є підтвердженням кількісної ефективності запропонованих заходів після їх інженерного впровадження, що заявляється автором як предмет подальших досліджень.

Оцінка змісту та структури дисертації, її завершеності й дотримання принципів академічної доброчесності

Дисертація має загальний обсяг 185 сторінок, з яких 112 сторінок становить основний текст. Робота містить вступ, чотири розділи, висновки,

список із 93 використаних джерел і чотири додатки. Матеріал проілюстровано 9 рисунками та 17 таблицями.

У вступі обґрунтовано актуальність теми, сформульовано наукове завдання, мету, об’єкт, предмет і завдання дослідження. Наведено положення наукової новизни, практичне значення та відомості про особистий внесок здобувача.

В першому розділі проаналізовано ройові системи БпЛА, моделі геопросторового сприйняття, стандарти управління ризиками та підходи до безпечності автономних систем. На цій основі сформульовано вимоги до комплексного методу оцінювання.

В другому розділі запропоновано модель життєвого циклу з чотирма стадіями: селекція знань, аналіз знань, навчання та експлуатація моделі. Сформовано таксономію ризиків і процедуру пріоритизування за MAI. Описано профілі змодельованих суджень, правила Монте-Карло-моделювання, очікуваний і спостережуваний профілі та порядок ітеративного уточнення таксономії.

В третьому розділі наведено результати пріоритизування. Для кожного з чотирьох профілів сформовано по 200 наборів матриць, а до аналізу включено лише реалізації з коефіцієнтом узгодженості меншим за 0,1. Експериментальну перевірку просторово-геометричних помилок виконано на 30 аерофотозображеннях із використанням п’яти мультимодальних ШІ-систем. Симуляційна частина досліджує поширення локальних і частково корельованих помилок для топологій «зірка», «сітка» та «ієрархічна сітка».

У четвертому розділі сформульовано практичні рекомендації щодо організації процесу оцінювання, контролю невизначеності й конфабуляцій, а також обмеження каскадного поширення помилок. Висновки відповідають змісту розділів і відображають отримані результати та межі їх застосовності.

Робота є завершеним дослідженням. Текст викладено послідовно, а додатки містять протоколи та відомості, потрібні для відтворення основних обчислювальних процедур.

Використані джерела перелічені в списку літератури, запозичені положення супроводжено посиланнями. Здобувач окремо задекларував використання ШІ-інструменту ChatGPT для пошуку, первинного опрацювання літератури, а також граматичної і стилістичної перевірки. У дисертації зазначено, що наукові положення, метод, результати й висновки перевірені та доопрацьовані автором, який несе відповідальність за їх коректність.

Повнота викладення основних результатів у наукових публікаціях

За темою дисертації опубліковано чотири статті, у яких подано основні наукові результати: дві статті у фахових виданнях України категорії «Б» та дві

статті у виданнях, що індексуються міжнародною науково метричною базою Scopus, одна з яких – у виданні першого квартиля за класифікацією SCImago Journal and Country Rank. Апробацію результатів засвідчують чотири публікації у збірниках матеріалів наукових конференцій.

Тематика публікацій охоплює будову й управління ройовими системами, застосування штучного інтелекту, пріоритизування ризиків і оцінювання ризиків моделей геопросторового сприйняття. Особистий внесок здобувача у працях зі співавторами конкретизовано.

Зауваження та дискусійні положення

До зауважень та дискусійних моментів щодо дисертаційної роботи можна віднести наступне.

1. Запропонована таксономія ризиків, які виникають під час використання ШІ-моделей геопросторового сприйняття, орієнтована переважно на технічні чинники. Було б корисним детальніше розглянути впливи цілеспрямованого зловмисного походження, наприклад, навмисне спотворення інформації від датчиків та сигналів GNSS, компрометацію окремих агентів тощо.

2. Кількісне пріоритизування ризиків у другому розділі мало б більшу емпіричну обґрунтованість, якби здійснювалося не лише на основі змодельованих профілів експертних суджень, але також шляхом залучення реальної групи експертів.

3. Було би доцільним довести дослідження кількісного зв'язку між рівнем ризику локальної помилки ШІ-моделі та конкретними показниками деградації колективної поведінки рою до стадії отримання можливості кількісного прогнозування системних наслідків реалізації таких ризиків.

4. Перевірка ризиків, які входять до запропонованої таксономії, здійснена нерівноцінно, тому було б доцільно більш чітко розмежувати експериментально підтверджені ризики, ризики, перевірені шляхом моделювання, та ризики, для яких запропоновано лише методику перевірки.

5. В тексті міститься певна кількість одруківок та помилок оформлення, зокрема:

– незрозуміле посилання "(див. 1.1)" на стр. 50;

– "загублений" маркер списку на початку стр. 101;

– невдалий перенос рядків в деяких таблицях,

а ілюстративний матеріал на сторінках 95–97 та 111 просто вставлений в текст і не оформлений як рисунки або таблиці.

Але згадані зауваження не є критичними та не зменшують загальну позитивну оцінку наукової та практичної значимості результатів дисертаційної роботи та рівня кваліфікації її автора.

Загальний висновок про дисертаційну роботу

Вважаю, що дисертаційна робота Кравчука Олександра Анатолійовича «Метод оцінювання ризиків життєвого циклу моделей штучного інтелекту геопросторового сприйняття для інформаційних підсистем роїв безпілотних літальних апаратів» є завершеним самостійним науковим дослідженням, в якому розв'язано актуальне наукове завдання, містить обґрунтовані положення наукової новизни та результати, що мають практичне значення для оцінювання ризиків інформаційних систем зі штучним інтелектом. За змістом, структурою, обґрунтованістю результатів, повнотою їх оприлюднення та рівнем академічної доброчесності дисертаційна робота 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 р. № 44, а здобувач Кравчук Олександр Анатолійович заслуговує на присудження ступеня доктора філософії в галузі знань 12 – Інформаційні технології за спеціальністю 122 – Комп'ютерні науки.

Рецензент:
Завідувач відділу математичного
та економетричного моделювання
Інституту проблем моделювання в
енергетиці ім. Г.Є. Пухова НАН України,
докт. техн. наук, ст. наук. співр. Сергій ГІЛЬГУРТ