

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0826U004512

Особливі позначки: відкрита

Дата реєстрації: 11-08-2026

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу: № НСВС/66/26 від 21.09.2026



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Вохранов Ілля Анатолійович

2. Illia Vokhranov

Кваліфікація:

Ідентифікатор ORCID ID: 0009-0000-1702-0460

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 03-09-2026

Спеціальність за освітою: Комп'ютерні науки

Місце роботи здобувача: ТОВАРИСТВО З ОБМЕЖЕНОЮ ВІДПОВІДАЛЬНІСТЮ "СТРІМЛАЙН ТЕХНОЛОГІЇ"

Код за ЄДРПОУ: 45587976

Місцезнаходження: вул. Бойчука Михайла, будинок 18, Київ, 01103, Україна

Форма власності: Приватна/недержавна

Сфера управління:

Ідентифікатор ROR:

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 16328

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 28.23.20, 28.23.37, 50.41.17

Тема дисертації:

1. Метод статичного аналізу вихідного коду програм із використанням великих мовних моделей
2. Method of Static Analysis of Program Source Code Using Large Language Models

Реферат/анотація:

1. Об'єкт дослідження — процеси статичного аналізу програмного коду, спрямовані на виявлення дефектів та вразливостей; предмет — нейросимвольні методи аналізу, що поєднують семантичне розуміння великих мовних моделей (LLM) з незалежною символьною верифікацією. Мета — розробка, формалізований опис та експериментальна оцінка нейросимвольного методу, який забезпечує контрольовану поведінку LLM через незалежну символьну верифікацію генерованих тверджень і дозволяє досягти високої повноти виявлення дефектів при збереженні точності та передбачуваності результатів. Дослідження виконано в рамках ініціативної науково-дослідної роботи СП-01/2023 «Впровадження сервіс-орієнтованого підходу до реалізації процесів діджиталізації суспільства» (№ держреєстрації 0123U101333). Методи дослідження: аналіз і систематизація наукових джерел; формалізація архітектури та алгоритмів; моделювання взаємодії компонентів; обчислювальний експеримент на реальних даних; порівняльний аналіз з існуючими підходами;

абляційне дослідження внеску кожного компонента. Програмно-апаратна база: власний програмний прототип, що поєднує ансамбль інструментів статичного аналізу, зіставлення AST-патернів і taint-аналіз із зверненням до API моделей GPT-4 та GPT-5.2; набір даних PySStuBs. Наукова новизна. Вперше запропоновано нейросимвольний метод, у якому на першому етапі LLM безпосередньо генерує структуровані твердження про дефекти коду, а символьний аналіз їх верифікує, на відміну від існуючих методів (IRIS, LLMSA, MoCQ), де символьний компонент виконує первинне виявлення, а нейронний лише фільтрує чи доповнює його результати. Інвертований напрямок взаємодії Neural to Symbolic дозволяє скористатися здатністю моделі сприймати семантику коду за одночасного жорсткого контролю згенерованих нею тверджень і мінімальних ризиків галюцинації. Вперше запропоновано спосіб верифікації структурованих тверджень про дефекти Explain-Verify Gate, у якому атрибути твердження (тип дефекту, локалізація, пропозиція мінімально достатнього виправлення, обґрунтування) визначають вибір та інтерпретацію подальших символьних перевірок — зіставлення AST-патернів, taint-аналізу, перевірки правдоподібності виправлення. Це забезпечує детермінізм верифікації без повторних звернень до мовної моделі за обґрунтуванням кожного її рішення. Удосконалено процес прийняття рішень щодо дефектів завдяки механізму Evidence-Weighted Suppression, який зважає докази з кількох незалежних джерел (інструменти статичного аналізу, AST-патерни, показник впевненості базового аналізатора) і відновлює валідні виявлення, відхилені символьною верифікацією за відсутності відповідних патернів, але підтверджені достатніми доказами. Це дозволяє збалансувати точність і повноту аналізу. Практичні результати. Оцінка на 1000 збалансованих зразках набору PySStuBs показала виявлення 379 дефектів проти 132 у базового LLM-підходу та 8 в ансамблі інструментів на основі правил, що відповідає зростанню recall на 187% та F1-score на 76,4% за співставного рівня точності (близько 0,51). Валідація з моделлю GPT-5.2 (n=500) підтвердила збереження переваг при переході на новіше покоління моделі (+132% F1-score, +363% recall), абляційне дослідження — стійкість архітектури до модифікації окремих компонентів (F1-score у діапазоні 0,654–0,668). Метод не потребує донавчання моделей і збору спеціалізованих наборів даних, допускає заміну нейронного компонента без модифікації решти системи, а покриття символьної верифікації розширюється додаванням правил, AST-патернів і taint-специфікацій для нових мов програмування та класів дефектів. Результат аналізу подається структурованим описом дефекту (тип, локалізація, обґрунтування, мінімально необхідне виправлення) разом із результатами символьної верифікації, що скорочує час розробника на опрацювання виявлень. Ступінь упровадження: розроблено та експериментально апробовано програмний прототип методу; результати використано для покращення інструментарію забезпечення якості програмного забезпечення в рамках виконання зазначеної науково-дослідної роботи. Упровадження методу в автоматизовані процеси контролю якості програмного коду планується. Галузь використання: інженерія програмного забезпечення, забезпечення якості та надійності програмних систем, аналіз захищеності вихідного коду, автоматизовані процеси контролю якості (CI/CD, code review, IDE), а також подальші дослідження нейросимвольних систем.

2. The object of the research is the processes of static analysis of program source code aimed at detecting defects and vulnerabilities; the subject is neurosymbolic analysis methods that combine the semantic understanding of large language models (LLMs) with independent symbolic verification. The aim is the development, formal description, and experimental evaluation of a neurosymbolic method that ensures controlled LLM behavior through independent symbolic verification of generated claims and enables high defect detection recall while preserving the precision and predictability of results. The research was carried out within the initiative research project SP-01/2023 "Implementation of a service-oriented approach to the realization of the processes of digitalization of society" (state registration No. 0123U101333). Research methods: analysis and systematization of scientific sources; formalization of the architecture and algorithms; modeling of component interactions; computational experiment on real data; comparative analysis with existing approaches; ablation study of the contribution of each component. Software and hardware base: an in-house software prototype combining an ensemble of static analysis tools, AST pattern matching, and taint analysis with queries to the APIs of the GPT-4 and GPT-5.2 models; the PySStuBs dataset. Scientific novelty. For the first time, a neurosymbolic method is

proposed in which the LLM at the first stage directly generates structured claims about code defects, and symbolic analysis verifies them, as opposed to existing methods (IRIS, LLMSA, MoCQ), where the symbolic component performs primary detection while the neural component merely filters or supplements its results. The inverted interaction direction Neural to Symbolic makes it possible to leverage the model's ability to perceive code semantics while maintaining strict control over the claims it generates and minimizing the risk of hallucination. For the first time, an approach to the verification of structured defect claims, the Explain-Verify Gate, is proposed, in which the attributes of a claim (defect type, location, proposed minimally sufficient fix, rationale) determine the selection and interpretation of subsequent symbolic checks: AST pattern matching, taint analysis, and fix plausibility checking. This ensures verification determinism without re-querying the language model for justification of each of its decisions. The defect decision-making process is improved through the Evidence-Weighted Suppression mechanism, which weighs evidence from several independent sources (static analysis tools, AST patterns, base analyzer confidence score) and recovers valid detections rejected by symbolic verification due to the absence of matching patterns but supported by sufficient evidence. This makes it possible to balance the precision and recall of the analysis. Practical results. Evaluation on 1,000 balanced samples of the PySStuBs dataset showed the detection of 379 defects versus 132 for the baseline LLM approach and 8 for the rule-based tool ensemble, corresponding to a recall improvement of 187% and an F1-score improvement of 76.4% at a comparable precision level (approximately 0.51). Validation with the GPT-5.2 model (n=500) confirmed that the advantages are preserved when moving to a newer model generation (+132% F1-score, +363% recall), and the ablation study confirmed the resilience of the architecture to modifications of individual components (F1-score within the range 0.654-0.668). The method requires neither model fine-tuning nor the collection of specialized datasets, allows replacement of the neural component without modifying the rest of the system, and its symbolic verification coverage is extended by adding rules, AST patterns, and taint specifications for new programming languages and defect classes. The analysis result is presented as a structured defect description (type, location, rationale, minimally required fix) together with symbolic verification results, which reduces the developer's time spent on processing detections. Degree of implementation: a software prototype of the method has been developed and experimentally validated; the results have been used to improve software quality assurance tooling within the framework of the aforementioned research project. Implementation of the method in automated code quality control processes is planned. Field of application: software engineering, quality and reliability assurance of software systems, source code security analysis, automated quality control processes (CI/CD, code review, IDE), as well as further research on neurosymbolic systems.

Державний реєстраційний номер ДіР: 0123U101333

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- Vokhranov, I., & Bulakh, B. (2026). VERIGATE: A verification gate-based neurosymbolic method for static code analysis. *Наука і Техніка Сьогодні*, 2(56), 1488–1507. - DOI 10.52058/2786-6025-2026-2(56)-1488-1507.
- Vokhranov, I., & Bulakh, B. (2024). Transformer-based models application for bug detection in source code. *Technology Audit and Production Reserves*, 5(2(79)), 6–15. - DOI 10.15587/2706-5448.2024.310822.
- Вохранов, І., & Булах, Б. (2023). Комбіновані підходи до статичного аналізу коду з використанням нейронних мереж. *Інфокомунікаційні та комп'ютерні технології*, 1(05), 183–193. - DOI 10.36994/2788-5518-2023-01-05-20.
- Petrenko, A. I., & Vokhranov, I. A. (2023). Neural networks: Studying their decision-making rules. *System Research and Information Technologies*, 2023(2), 74–85. - DOI 10.20535/SRIT.2308-8893.2023.2.06.

- Вохранов, І. (2025). Нейросимвольні підходи в аналізі програмного коду: огляд сучасного стану та перспективи розвитку. 4th International Scientific and Practical Conference «Science and Information Technologies in the Modern World», 161–164.

Наукова (науково-технічна) продукція: методи, теорії, гіпотези

Соціально-економічна спрямованість: підвищення продуктивності праці; підвищення автоматизації виробничих процесів

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Планується до впровадження

Зв'язок з науковими темами: 0123U101333

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Булах Богдан Вікторович

2. Bohdan V. Bulakh

Кваліфікація: к. т. н., доц., 05.13.12

Ідентифікатор ORCID ID: 0000-0001-5880-6101

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Сенів Максим Михайлович

2. Maksym M. Seniv

Кваліфікація: к. т. н., доц., 01.05.03

Ідентифікатор ORCID ID: 0000-0003-1044-4628

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Глибовець Андрій Миколайович

2. Andrii Glybovets

Кваліфікація: д. т. н., професор, 01.05.03

Ідентифікатор ORCID ID: 0000-0003-4282-481X

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Стіренко Сергій Григорович

2. Serhii H. Stirenko

Кваліфікація: д. т. н., професор, 05.13.05

Ідентифікатор ORCID ID: 0000-0001-5478-0450

Додаткова інформація: <https://scholar.google.com.ua/citations?user=abSxLMIAAAAJ&hl=uk&oi=ao>;
<https://www.scopus.com/authid/detail.uri?authorId=54421204800>

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Харченко Костянтин Васильович

2. Kostiantyn V. Kharchenko

Кваліфікація: к. т. н., 05.13.12

Ідентифікатор ORCID ID: 0000-0002-7334-8038

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

VIII. Заключні відомості

**Власне Прізвище Ім'я По-батькові
голови ради**

Чумаченко Олена Іллівна

**Власне Прізвище Ім'я По-батькові
головуючого на засіданні**

Ільченко Михайло Юхимович

**Відповідальний за підготовку
облікових документів**

Вохранов Ілля Анатолійович

Реєстратор

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є
відповідальним за реєстрацію наукової
діяльності**



Юрченко Тетяна Анатоліївна