Отзыв

на автореферат диссертации Романова Александра Сергеевича

«Методология идентификации автора текстовой информации для решения задач кибербезопасности», представленную на соискание ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность

В условиях глобальной цифровизации и расширения коммуникационных возможностей вопросы установления авторства текстовых сообщений и исходных кодов программ приобретают особую значимость, поскольку напрямую связаны с обеспечением государственной, общественной и корпоративной безопасности. Разработка научно обоснованной комплексной методологий идентификации автора становится неотъемлемой частью противодействия угрозам, возникающим в киберпространстве.

Актуальность работы не вызывает сомнений и обусловлена необходимостью эффективного противодействия угрозам, связанным с анонимным распространением деструктивного, экстремистского и фейкового контента, а также с расследованием инцидентов в киберпространстве. Представленная диссертация отвечает современным вызовам и соответствует приоритетным направлениям развития информационной безопасности.

Развитие методов искусственного интеллекта и машинного обучения открывает новые возможности для решения подобных задач, что убедительно доказывает диссертационное исследование А.С. Романова.

Научная новизна исследования подтверждается рядом оригинальных результатов, полученных автором:

1. Разработана комплексная методология идентификации автора текста, устойчивая к атакам и унифицированная для обработки как естественно-языковых текстов, так и программного кода.

2. Создана модель написания текста в киберпространстве, интегрирующая семантические признаки всех уровней, атрибуты автора и специфику цифровой среды.

3. Создана методика идентификации автора текста на основе гибридной архитектуры GRU+CNN и SVM с оптимизацией признакового пространства, адаптированная для открытого/закрытого сценария и AI-текстов.

4. Разработана методика атрибуции исходного кода на основе глубокой трансформерной модели CodeBERT, обеспечивающая идентификацию автора в современных сценариях разработки ПО.

5. Предложена методика возрастной классификации автора с использованием fastText и методов компьютерного зрения.

6. Создана методика детекции деструктивных и экстремистских текстов и их авторов с применением ансамбля моделей (GRU+CNN, BERT) и методов трансферного обучения, ориентированная на соответствие требованиям законодательства РФ.

7. Разработана методика определения пола и гендера (включая ЛГБТ+) автора русскоязычного текста с использованием ансамбля SVM, CNN на триграммах Катца и BERT.

8. Предложена методика проверки однородности текста и обнаружения заимствований на основе сиамских нейросетей для обнаружения заимствований в условиях открытого множества авторов и использования AI-генерации.

Основные практические результаты включают:

1. Разработку репрезентативных корпусов текстовых данных, содержащих художественные, любительские, научные тексты, сообщения из социальных сетей и исходные коды программ с верифицированными атрибутами авторов, что создает методологическую базу для перспективных исследований в области компьютерной лингвистики и информационной безопасности.

2. Методика идентификации автора естественного текста показала впечатляющие результаты точности - 99% для 2-5 авторов, 94% для 10 авторов и 88% для 20 авторов. При работе с короткими текстами (до 250 символов) точность достигает 96,3%.

3. В методике идентификации текстов деструктивной направленности получен прирост точности до 10% при идентификации автора, до 12% при определении деструктивного контента. Совокупный прирост до 20% при комплексном подходе.

4. Методика определения пола и гендера продемонстрировала 88-92% точности при разграничении мужчин и женщин, 93% точности при определении признаков ЛГБТ-сообщества и 68% точности при классификации шести гендерных групп.

5. Методика определения возраста автора текста обеспечила 82% точности для двух возрастных категорий, 63% точности для трех возрастных категорий.

6. Методика идентификации автора исходного кода программы демонстрирует 93% точности в простых случаях и до 85% точности в современных сценариях разработки ПО.

7. Методика проверки однородности текста показала 90% и более точности при определении авторства фрагментов.

Апробация на 26 конференциях и публикации в высокорейтинговых изданиях (18 публикаций в изданиях ВАК, 13 в Web of Science/Scopus), а также использование результатов при выполнении 7 грантов и хозяйственных договоров подтверждают достоверность и обоснованность полученных решений.

Вместе с тем, к незначительным замечаниям можно отнести:

1. В работе недостаточно подробно описаны перспективы масштабирования предложенных методик на языки, отличные от русского, хотя потенциал для такого развития очевиден.

2. На рис. 9 стр. 27 приведены результаты идентификации автора исходного кода для 13 языков программирования (C++, Java, JS, …). Почему были выбраны именно эти языки?

3. В представленном автореферате есть погрешности оформления: название рис. 8 смещено на стр. 27; заголовок и первая строка табл. 6 находятся на стр. 33, тогда как остальные строки на стр. 34; последняя строка табл. 7 смещена на стр. 37; источник 45 сместился на стр. 48. в виде висячей строки. Погрешности явно связаны с технической оплошностью, потому что сам автореферат оформлен на высоком уровне и строго по ГОСТ.

Работа А.С. Романова полностью соответствует паспорту специальности 2.3.6, вносит существенный вклад в развитие методологии информационной безопасности и решает крупную научную проблему, имеющую важное значение для отрасли – тем самым соответствует критериям, установленным в п. 9-14 «Положении о порядке присуждении ученых степеней» ВАК Российской Федерации, утвержденного постановлением Правительства Российской Федерации №842 (редакция от 16.10.2024). Автореферат адекватно отражает основное содержание диссертации и соответствует п. 25 «Положения о порядке присуждении ученых степеней».

Считаю, что Романов Александр Сергеевич достоин присуждения ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Я, Сухопаров Михаил Евгеньевич, даю свое согласие на обработку персональных данных и на их включение в документы, связанные с работой диссертационного совета.

Заместитель директора по научной работе
Санкт-Петербургского филиала АО «НПК «ТРИСТАН»

доктор технических наук, доцент Сухопаров М.Е.

05.12.13 – «Системы, сети и устройства телекоммуникаций»

Подпись Сухопарова Михаила Евгеньевича заверяю:

Заместитель директора по качеству
СПбФ АО «НПК «ТРИСТАН» Кротов А.В.

«22» сентября 2025 г.

СПбФ АО «НПК «ТРИСТАН»
ИНН 7718213897 КПП 780443001
Адрес: 195220, г. Санкт-Петербург, пр. Непокоренных, д.47
тел.: +7 (901) 9707457
e-mail: mail@spb3stan.ru