ОТЗЫВ

официального оппонента Спицына Владимира Григорьевича

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

Актуальность выбранной темы

Затронутые в диссертации задачи бесспорно являются актуальными.

Нарушение авторских и смежных прав на текстовые произведения, выявления информации, пропагандирующей нетрадиционные сексуальные отношения, педофилию и смену пола, материалов, содержащих подстрекания к насилию, террористическим действиям или участию в экстремистских организациях, недопущение детей и подростков до контента, который имеет возрастное ограничение, определение авторов вирусных программ, отвечающих за массовые сбои критической инфраструктуры РФ, или выявление факта неправомерного использования чужого программного кода в коммерческих продуктах являются актуальными и острыми вызовами научному и техническому сообществу.

Комплексное решение подобных вопросов невозможно без создания методологии идентификации автора текста и таких характеристик автора как пол, возраст, настроение, идеологические взгляды и др. Диссертационное исследование Романова А.С. с этой точки зрения создает основу для их решения, предлагая актуальное методическое, алгоритмическое и программное обеспечение обработки текстов для решения задач кибербезопасности.

Основное содержание диссертационной работы

Диссертационная работа имеет объем 400 страниц и состоит из введения, 6 глав, заключения, списка литературы и 4 приложений. Библиография содержит 334 русскоязычных и англоязычных источника.

Изложение полученных автором результатов последовательно и логично, диссертация носит завершенный характер.

В **первой главе** проводится анализ работ по теме диссертации. Рассмотрены работы А.С. Сурковой, А.Р. Дубовик, Ю.Н. Орловой, А.Н. Ульченко, О.М. Атаевой, Т.П. Соколовой, С.В. Чащина, А.С. Коляды, В.Д. Гогунского, Н.Д. Москина, А.К. Ковалёва, И. Маркова, И.В. Огорелкова, А.Б. Хазовой, Д.А. Галкиной, А.Г. Сбоева, А.А. Воробьевой, М.Е. Сухопарова, В.Д. Стремоухова, А.О. Корней, Ю.В. Рубцовой, С.В. Вычегжанина, Т.А. Литвиновой, В.А. Минаева, А.А. Гончарова, Ю.В. Давыдовой, А.О. Исхаковой, Д.Н. Буинцева, A. Abbasi, S. Ishihara, F.

Johansson, H. Gomez-Adorno, S. Hedegaard, J. G. Simonsen, F. Rangel, A. Karami, B. Alsulami, A. Caliskan-Islam, S. Afroz, M. Abuhamad, E. Quiring и др. ученых, занимавшихся вопросами идентификации автора текста и профилирования текстов. На основе анализа выделяются актуальные задачи кибербезопасности, связанные с анализом текста. Подчеркивается необходимость систематизации и модернизации существующих подходов таким образом, чтобы они учитывали особенности русского языка, семантику и контекст, возможные способы атаки, и тот факт, что искусственно-языковые тексты требуют применения принципиально иных методов для анализа и определения авторства, чем естественные.

Вторая глава является теоретической. В ней предлагается универсальная методология идентификации автора для решения задач кибербезопасности, связанных с классификацией текстов. Методология основана на двух базовых методиках идентификации автора естественного и искусственного текста, а также модели создания текста автором в киберсреде, учитывающей анализ творческой, повседневной и профессиональной деятельности и порождаемой при этом текстовой информации. Отмечается, что использование глубокого обучения предпочтительно для поиска явных и неявных информативных признаков текста.

В **третьей главе** приводится математическая постановка задач собственно идентификации автора текста в нескольких случаях:

1) когда предполагаемый автор известен и есть образцы его текстов (закрытая атрибуция);

2) когда нужно определить написаны ли тексты одним и тем же человеком (верификация);

3) когда человек, реально написавший текст, не рассматривается в качестве предполагаемого автора, но необходимо определить, что ни один из авторов, тексты которых имеются, истинным авторов не является (открытая атрибуция).

Далее проводится серия вычислительно сложных экспериментов, направленных на установление комбинации классификаторов, с помощью которой можно однозначно идентифицировать автора текста. Итоговая методика на основе архитектур GRU и CNN достигает точности 94,2%, 98,2% и 93,5% для художественных, любительских и коротких текстов, соответственно.

Четвертая глава описывает аналогичную серию экспериментов для текстов программ для популярных языков программирования. Автором предлагается методика на основе архитектуры CodeBERT, модифицированной несколькими классификационными слоями. Полученная точность 93%, позволяет сделать вывод о её высокой эффективности в сравнении с аналогами.

В **пятой главе** на основе базовой методики идентификации автора текста, разработаны методики:

1) идентификации возраста автора, в том числе учитывается возрастная категория несовершеннолетних людей;

2) идентификации пола и гендера, при этом используются в том числе тексты представителей запрещенного в РФ движения ЛГБТ;

3) идентификации текстов и авторов экстремистских движений и иностранных агентов;

4) проверки научных текстов на однородность и плагиат, учитывающая сгенерированные тексты и случай «открытой атрибуции».

В методиках используются семантические особенности текстов и специфические подходы машинного обучения, учитывающие уникальный характер каждой задачи. В частности, интересным представляется использование трансферного обучения, сиамских нейронных сетей и методов компьютерного зрения в сочетании с привычными методами машинного обучения, позволившее добиться точности более 90% при решении каждой из задач.

В **шестой главе** приводится описание разработанных программных продуктов, алгоритмов и баз данных для решения исследовательских и прикладных задач информационной безопасности в области анализа текста. Стоит отметить, что разработанные методики доведены до готовых программных продуктов, программы «PyAuthorship», «Авторовед», «Age Detector», «CoDEtective», «Moodorrude», «PyDestructiveDetector», «ProGender» и базы данных зарегистрированы в Роспатенте РФ и используются в деятельности предприятий различной формы собственности, что подтверждено соответствующими свидетельствами о регистрации и актами внедрения.

Таким образом, можно сделать вывод о том, что задачи рассматриваются автором последовательно, в порядке, необходимом для достижения поставленной в диссертации цели.

Научная новизна

В результате анализа текста диссертации и опубликованных научных работ Романова А.С. можно заключить, что к основным научным результатам, обладающим новизной, следует отнести:

1. Методологию идентификации автора текста, учитывающую особенности естественно- и искусственно-языковых текстов и возможные атаки на методы идентификации.
2. Авторскую модель формирования текста в киберсреде, включающую анализ семантики и информативных признаков на разных уровнях, а также учитывающую специфику среды, личностные атрибуты автора и характер текстообразующей деятельности.
3. Методику идентификации автора текста, использующую гибрид GRU+CNN и SVM с отбором ключевых признаков через генетический алгоритм и

регуляризацию. Новизна подхода заключается в учете случаев открытой/закрытой атрибуции и анализа текстов, созданных ИИ.

4. Метод установления авторства программного кода на основе классификатора CodeBERT. Особенность методики – анализ сложных ситуаций, включая запутывающие преобразования, совместную разработку, применение код-стайлов и кодов, сгенерированных ИИ.

5. Методику классификации текстов на две и более возрастные группы, в котором применяется fastText, особенностью является использование алгоритмов компьютерного зрения для очистки исходных данных перед обучением.

6. Методику идентификации запрещенных текстов, имеющих экстремистский характер или признаки создания иноагентами, а также определения автора таких текстов. Особенностью является использование семантических признаков текста, трансферного обучения и гибрида GRU и CNN.

7. Ансамбль SVM, CNN и BERT, обученный на сглаженных частотах триграмм и семантических признаках, для определения пола и гендерной идентичности автора текста. Методика впервые учитывает гендеры ЛГБТ для русского языка.

8. Инновационную методику детекции заимствований и оценки текстовой однородности, использующую сиамские архитектуры нейронных сетей. В отличие от существующих решений, предложенный подход обеспечивает эффективное распознавание заимствований в сценариях с неограниченным множеством потенциальных авторов, включая случаи применения генеративных моделей.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность

Обоснованность и достоверность полученных результатов и выводов сомнений не вызывают. Соискатель корректно применяет аппарат искусственных нейронных сетей, математической статистики, методы исследований, проводит эксперименты на объемных текстовых корпусах из различных источников, использует для проверки общеупотребимые метрики оценки качества классификации и кластеризации, соотносит их с результатами других научных коллективов. Внедрение результатов в практику АО «Национальный Инновационный Центр», ООО «СИБ», ООО «НТР», ООО «Сибэдж», войсковой части 51952, ИШИТР ТПУ, ОБК УМВД России по Томской области, экономического факультета МГУ имени М.В. Ломоносова, Томского государственного университета систем управления и радиоэлектроники показывает положительный эффект.

Теоретическая и практическая значимость

В работе предложена новая методология идентификации автора текста, которая систематизирует, развивает и дополняет теоретические основы создания математического, алгоритмического и программного обеспечения решения задач

информационной безопасности, связанных с анализом индивидуально-личностных характеристик автора.

Практическая значимость работы Романова А.С. состоит в том, что предложенные методики работают точнее, быстрее, требуют меньшего количества текста для анализа и учитывают случаи, которые ранее не рассматривались для русского языка. В частности, методика идентификации автора естественно-языкового текста работает на 2–14% точнее, чем аналоги; идентификации автора исходного кода – на 20-30% точнее. А созданные на их основе методики определения запрещенных законодательством РФ текстов позволяют определить человека, который их написал, на 30% точнее.

Апробация и публикации

Содержание диссертационной работы раскрыто в публикациях соискателя и прошло необходимую апробацию.

Основные положения опубликованы в 94 научных работах, в том числе 18 статей в журналах, рекомендованных ВАК РФ, 13 статей включены в реферативные базы данных WOS и Scopus. Результаты работы отражены в монографии, получено 8 свидетельств Роспатента РФ о регистрации программ для ЭВМ и 3 свидетельств о регистрации баз данных.

Замечания

1. Почему в задаче идентификации авторства текста для отбора признаков были выбраны методы, основанные на генетических алгоритмах, SHAP и регуляризации, вместо использования одного из множества метаэвристических методов, основанных на поведении популяций животных, которые успешно применяются для решения широкого спектра задач?

2. Известно, что ChatGPT «галлюцинирует» при генерации текстов. Учитывалось ли это при составлении датасета?

3. Каким образом обеспечивался сбор данных экстремистского характера, ведь такой контент активно блокируется в мессенджерах и социальных сетях?

4. Каким образом подбиралось количество эпох при обучении для каждой из моделей?

5. Объемные таблицы (например, 3.4 и 5.14) стоило вынести в приложение.

Отмеченные недостатки не снижают качества исследования и не влияют на полученные научные и практические результаты. Общая оценка работы – положительная.

Заключение

Диссертационная работа А.С. Романова «Методология идентификации автора текстовой информации для решения задач кибербезопасности» является

законченным научным исследованием и соответствует паспорту специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность (пп. 1, 5, 12, 13). Диссертация написана автором самостоятельно, содержит новые научные результаты и положения. Автореферат полностью отражает содержание диссертации.

Романовым А.С. решена важная научная проблема по созданию и развитию методов идентификации автора печатного текста, имеющая важное хозяйственное и техническое значение. Внедрение полученных автором результатов вносит значительный вклад в развитие основ для создания программных систем и новых безопасных информационных технологий нашей страны.

Считаю, что диссертационная работа соответствует требованиям пп. 9-14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 г. № 842, предъявляемым к докторским диссертациям. Ее автор, Романов Александр Сергеевич, заслуживает присуждения ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Я, Спицын Владимир Григорьевич, даю свое согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

Официальный оппонент,
доктор технических наук, профессор, профессор отделения информационных технологий ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»,
634050, г. Томск, проспект Ленина, 30.
Email: spvg@tpu.ru

Владимир Григорьевич Спицын
«01» сентября 2025

Диссертация на соискание ученой степени доктора технических наук защищена по специальностям 05.13.16 – Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях, 01.04.03 – Радиофизика (2000 год).

Подпись Спицына Владимира Григорьевича удостоверяю

И.О. ученого секретаря НИ ТПУ

Валерия Дмитриевна Новикова
«01» сентября 2025