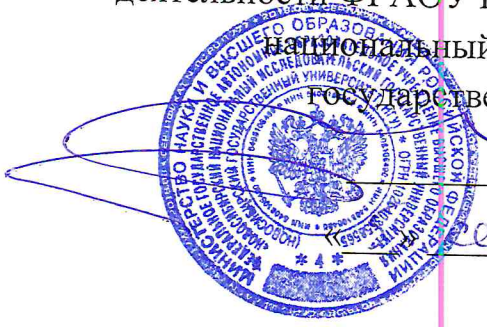
УТЕРЖДАЮ

Проректор по научно-исследовательской деятельности ФГАОУ ВО «Новосибирский национальный исследовательский государственный университет»



_____ Д.В.Чуркин

_____ 2025 г.

ОТЗЫВ

ведущей организации, Федерального государственного автономного образовательного учреждения высшего образования «Новосибирский национальный исследовательский государственный университет», на диссертационную работу 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

Актуальность темы исследования

В диссертационной работе Романова А.С. решается важная научная проблема – идентификация авторства текстов. Проблема является **актуальной** в контексте информационной безопасности по нескольким причинам.

Во-первых, с увеличением объемов информации и распространением цифровых технологий возрастает риск распространения дезинформации и фальсификаций. Анонимные источники могут манипулировать общественным мнением, используя поддельные документы или публикации. Установление подлинности автора текста позволяет выявить подобные попытки манипуляции и защитить пользователей от ложной информации.

Во-вторых, идентификация авторства играет ключевую роль в юридических и правовых аспектах. В случае споров о плагиате, нарушении авторских прав или клевете важно точно определить, кто является автором конкретного текста. Это позволяет защитить права интеллектуальной собственности и предотвратить злоупотребления, что особенно актуально в эпоху, когда контент может быстро распространяться через интернет.

Кроме того, в корпоративной среде идентификация авторов текстов важна для защиты внутренней информации и предотвращения утечек данных. Знание того,

кто создает и распространяет информацию, помогает организациям контролировать доступ к конфиденциальным данным и минимизировать риски, связанные с возможными угрозами от недобросовестных сотрудников или внешних злоумышленников.

Таким образом, идентификация автора текста становится неотъемлемой частью системы информационной безопасности, способствуя обеспечению достоверности информации, защите прав интеллектуальной собственности и минимизации рисков утечек данных.

Общая характеристика содержания диссертационной работы

Диссертационная работа изложена на 400 страницах, состоит из введения, шести глав, заключения, списка литературы из 334 наименований и четырех приложений. Оформление и основное содержание работы соответствует рекомендациям ВАК и ГОСТ.

Во введении обоснована актуальность темы, сформулированы проблема, цель и задачи исследования, приведены научная новизна, теоретическая и практическая значимость, а также представлены основные защищаемые положения.

Первая глава посвящена ключевым задачам анализа текста определения авторства текста и существующим способам их решения. На основе глобальной задачи определения авторства сформулирован ряд прикладных подзадач кибербезопасности, требующих обособленных подходов к решению. Для каждой задачи выполнен обзор существующих решений, приведен сравнительный анализ, выделены преимущества и недостатки. Данная глава определила цель работы – создание методологии идентификации автора текстовой информации, включая естественно-языковые тексты и исходные коды программ, для решения задач кибербезопасности.

Во **второй главе** предложена методология идентификации авторства текстовой информации для решения задач кибербезопасности. Дано подробное описание, включая математический аппарат, основных составляющих данной методологии: модели создания автором текста в киберсреде, модели представления текстовой информации, алгоритмов анализа текста, методов принятия решений, оптимизации, сглаживания, снижения размерности и отбора признаков, а также описание атак на методы идентификации авторства.

Важно отметить, что помимо собственно авторства текста, А.С. Романовым предлагаются методы определения авторских признаков таких как пол, возраст, настроение и предлагается комплексная методология идентификации автора текста, чего не делали его предшественники.

Главы 3–5 содержат основные численные результаты апробации разработанных методик и составляющих их моделей и алгоритмов.

Третья глава посвящена методике идентификации автора естественно-языкового текста и результатам ее апробации. Разработанная методика позволяет учитывать случаи закрытой атрибуции автора, подразумевающей наличие истинного автора текста среди возможных авторов, открытой атрибуции текста, в рамках которой истинный автор может отсутствовать, а также верификации автора.

В **четвертой главе** представлена методика идентификации автора искусственно-языкового текста. Методика направлена на решение задачи определения автора исходного кода в простых и сложных случаях, где под сложными подразумевается наличие таких осложняющих факторов как обфускация, командная разработка, следование стандартам кодирования и искусственная генерация исходного кода.

Пятая глава описывает примеры применения методологии идентификации авторства к решению важных народно-хозяйственных задач: определение текстов экстремистской и деструктивной направленности, определение возраста автора текста, определение пола и гендерной принадлежности автора текста, определение однородности и поиска заимствований текста.

Шестая глава посвящена алгоритмическому и программному обеспечению, реализующему методики идентификации авторства естественно- и искусственно-языковых текстов, а также наборам данных для анализа текстов в рамках решения задач кибербезопасности.

В **заключении** представлены основные результаты и выводы по работе. Поставленные автором диссертационной работы задачи выполнены в полном объеме, обозначенная цель достигнута.

Список литературы содержит обширную библиографию, в которой содержится достаточное количество актуальных источников по тематике диссертационного исследования.

Научная новизна проведенных исследований и полученных результатов

Новизна теоретических и практических результатов, полученных автором:

1. Разработана комплексная методология идентификации авторства текстов, которая учитывает особенности текстов не только написанных человеком, но и искусственно созданных, а также возможные атаки на методы идентификации, что делает её более универсальной и устойчивой к современным вызовам.

2. Предложена новая модель создания текста автором в киберсреде, которая впервые интегрирует семантические особенности текста, информативные признаки на различных уровнях анализа, специфику цифровой среды, а также атрибуты автора и характер его деятельности.

3. Разработана методика идентификации авторства естественных текстов, основанная на комбинации рекуррентных и сверточных нейросетей, метода опорных векторов, с использованием генетического алгоритма для отбора признаков и методов регуляризации. Методика впервые охватывает случаи открытой и закрытой атрибуции, а также тексты, созданные генеративными нейросетями.

4. Разработана методика идентификации авторства исходного кода программ с использованием глубокой модели CodeBERT. Отличительной особенностью методики является учет случаев обфускации, генеративных нейросетей, командной разработки, использования нескольких языков программирования, стандартов оформления исходных кодов.

5. Представлена методика классификации по возрастным группам на основе текста, отличающаяся использованием комбинации методов компьютерного зрения для фильтрации недостоверных обучающих данных и модели fastText для бинарной и мультиклассовой классификации по возрасту.

6. Разработана методика определения текстов экстремистского и деструктивного характеров и идентификации авторства таких текстов. Методика отличается применением методов семантической кластеризации, моделей рекуррентных и сверточных нейросетей и BERT, а также трансферного обучения.

7. Представлена новая методика, направленная на определение пола и гендера автора текста на русском языке, которые впервые учитывает гендеры ЛГБТ. Методика отличается использованием ансамбля из метода опорных векторов, обученного на признаках, сформированных методами семантической кластеризации, сверточной нейросети, обученной на частотных распределениях триграмм, сглаженных по методу Катца, а также BERT.

8. Разработана методика проверки однородности текста и поиска заимствований, которая отличается применением сиамских нейросетей, обученных с тройной и контрастивной функциями потерь. Данная методика впервые решает задачу поиска заимствований для открытого множества вероятных авторов и для искусственной генерации текстов.

Обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации

В диссертации рассматривается весь комплекс вопросов, соответствующих научному исследованию. Обоснованность и достоверность полученных

результатов не вызывают сомнений и подтверждаются корректным использованием методов, обоснованными доказательствами научных положений, согласованностью полученных результатов с другими научными группами, результатами вычислительных экспериментов с программными реализациями разработанных моделей и алгоритмов, а также их апробацией на прикладных задачах в ФГАОУ ВО «ТУСУР», ООО «НТР Томск», ООО «СИБ», ООО «Сибэдж», ООО «НИЦ», ФГАОУ ВО НИ ТПУ, ОБК УМВД России по Томской области, войсковой части 51952 и на экономическом факультете МГУ имени М.В. Ломоносова, подтвержденной соответствующими документами.

Основные результаты исследований опубликованы в 18 статьях в журналах из перечня ВАК, 13 – в изданиях, индексируемых Web of Science/Scopus. Получены 8 свидетельств о регистрации программ для ЭВМ и 3 о регистрации БД. Результаты исследования докладывались на российских и международных конференциях.

Практическая и научная значимость полученных автором результатов

Практическая и научная значимость диссертационной работы состоит в том, что в результате ее выполнения решена важная научно-техническая проблема, заключающаяся в разработке научных основ идентификации автора текстовых материалов за счет применения современных методов машинного обучения. Решение проблемы имеет важное хозяйственное значение, поскольку вносит значительный вклад в развитие, прежде всего, информационной безопасности.

Теоретическая значимость работы заключается в разработке и обосновании новых подходов к идентификации авторства текстов, включая тексты деструктивного и экстремистского характера, с учетом современных проблем цифровой среды, таких как использование генеративных нейросетей, обфускация кода. Предложенные автором модели и методики расширяют существующие представления о лингвистических и семантических особенностях текстов, а также о влиянии цифровой среды на процесс их создания. Важным теоретическим вкладом является разработка новых методов анализа текстов, интегрирующих традиционные лингвистические подходы и современные технологии машинного обучения, включая глубокие нейронные сети. Особенно хотелось бы отметить, что разработанная А.С. Романовым методология учитывает не только обычные тексты, но и исходные коды программ.

Практическая значимость работы определяется возможностью использования предложенных методик и моделей для решения прикладных задач в различных областях. Разработанные подходы могут использоваться для защиты интеллектуальной собственности; идентификации автора сообщений в сети Интернет и продленной аутентификации пользователей социальных сетей; выявления признаков пропаганды нетрадиционных ценностей, деструктивной

информации, а также текстов экстремистской направленности, запрещенных законодательством Российской Федерации; определения автора-вирусописателя. Предложенные практико-ориентированные методики определения пола, возраста и других характеристик автора могут быть применены в правоохранительной практике и в рамках проведения различных криминалистических исследований. Также предложенные подходы открывают новые возможности для автоматизированного анализа текстов в условиях растущего объема данных, включая тексты, созданные генеративными моделями, что найдет свое применение в образовательной и академической среде для выявления плагиата.

Рекомендации по использованию результатов и выводов диссертационной работы

Теоретические результаты исследования, в частности методология идентификации автора естественного текста и исходных кодов, могут использоваться как теоретическая база для разработки новых методик определения признаков текста и характеристик автора, что весьма важно для решения задач кибербезопасности и научно-технологического развития Российской Федерации.

Обширное внедрение в деятельность компаний, указанных в диссертации, подтвержденное актами внедрения, показывает, что потребность в подобных инструментах есть в различных отраслях.

Практические результаты диссертационного исследования могут быть использованы в научных и научно-производственных организациях, занимающихся вопросами защиты информации и/или разработкой систем обеспечения информационной безопасности для идентификации и аутентификации пользователей по тексту и контент-анализа сетевого трафика.

Полученные результаты найдут свое применение в образовательном процессе для проверки студенческих работ на плагиат и повышения качества подготовки обучающихся.

Разработанные методика и программные средства для анализа исходных кодов программ могут использоваться в деятельности компаний, занимающих разработкой программного обеспечения, для код-ревью и проверки соответствия кода стандартам и передовым практикам программирования, а также для обеспечения цифрового суверенитета Российского Федерации.

Представленные в работе текстовые корпуса имеют особенную ценность и могут быть использованы для обучения нейросетевых моделей для решения задач не только в области информационной безопасности, но и в криминалистике, филологии и других областях.

Замечания и вопросы по диссертации

1. В диссертации не приводится определение понятия «киберсреда».
2. В тексте диссертационной работы присутствует большое количество аббревиатур. Следовало бы сделать раздел со списком сокращений и аббревиатур.
3. В первой главе приводится достаточно подробный анализ существующих работ в области анализа текстов, однако в автореферате эта информация отсутствует. Целесообразно было бы привести краткую сводку по этим работам в автореферате.
4. В аббревиатурах и соответствующих им понятиях встречаются разночтения. Например, в первой главе диссертации введена аббревиатура SVM для «метода опорных векторов», однако в шестой главе аналогичный метод называется «машиной опорных векторов».
5. Часть второй главы диссертационной работы носит описательный характер. Текст, посвященный различным методам анализа текста, можно сократить без существенного ущерба научному содержанию.
6. В таблицах с результатами экспериментов приведен также показатель временных затрат на обучение моделей. Однако из текста не ясно, является ли это временем обучения одной модели или совокупным временем обучения всех моделей для конкретного эксперимента?
7. В шестой главе следовало бы привести подробности технической реализации программного обеспечения. Какие программные решения/инструменты/библиотеки использовались для реализации алгоритмов машинного и глубокого обучения?
8. Не раскрыто, можно ли использовать разработанные в шестой главе программные решения как модули в составе других программных комплексов для решения смежных задач кибербезопасности? Все ли из них имеют графический интерфейс?

Отмеченные недостатки не носят принципиального характера и не влияют на общую положительную оценку диссертационной работы.

Заключение по работе

Диссертационная работа Романова Александра Сергеевича является самостоятельной и законченной научно-квалификационной работой, решающей научную проблему с важным народно-хозяйственным значением и вносящей значительный вклад в развитие методов анализа текстовой информации и информационной безопасности.

Работа написана понятным научным языком, имеет научную и практическую значимость. Представленные в ней результаты достоверны, выводы и заключения на основе проведенных исследований полны и обоснованы. Содержание

автореферата полностью соответствует тексту диссертационной работы. Все основные результаты диссертации опубликованы и доложены научному сообществу на научных мероприятиях.

Диссертационная работа содержит научно-квалификационные признаки, соответствующие требованиям пп. 9-14 «Положения о порядке присуждения ученых степеней» ВАК РФ, утвержденного постановлением Правительства РФ от 24.09.13 №842 (редакция от 16.10.2024), предъявляемых к докторским диссертациям, а Романов Александр Сергеевич заслуживает присуждения ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Диссертация, автореферат и представленный отзыв заслушаны, обсуждены и одобрены на заседании объединенного семинара кафедры математического моделирования Новосибирского государственного университета и отдела информационных технологий и проблем мониторинга Федерального исследовательского центра информационных и вычислительных технологий «Информационные технологии» (протокол № 1 от 2 сентября 2025 года).

Заведующий кафедрой
математического моделирования НГУ,
д.т.н., доцент Владимир Борисович Барахнин
4.09.2025

Сведения о ведущей организации

Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»
Адрес: 630090, Новосибирская область, г. Новосибирск, ул. Пирогова, д. 1
Тел. +7 (383) 363-40-00. Факс +7 (383) 330-42-80
Адрес в интернете: https://www.nsu.ru
E-mail: rector@nsu.ru

Подпись Владимира Борисовича Барахнина удостоверяю:

Ведущий специалист
управления кадров Наталья Николаевна Каврига

Подпись
Специалист Управления кадров НГУ
05 09 2025