На правах рукописи

Тетерин Николай Николаевич

Разработка методов и средств интеллектуальной обработки мультимодальных данных из сети Интернет

Специальность 2.3.8 —

«Информатика и информационные процессы»

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

Москва — 2025

Работа выполнена на кафедре «Математическая кибернетика и информационные технологии» ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ).

Научный руководитель: доктор технических наук, доцент
Саратова Татьяна Евгеньевна

Официальные оппоненты: **Преображенский Андрей Петрович,**
доктор технических наук, профессор, Воронежский институт высоких технологий – автономная некоммерческая образовательная организация высшего образования, научный клуб, руководитель

Немтинов Владимир Алексеевич,
доктор технических наук, профессор, федеральное государственное бюджетное образовательное учреждение высшего образования «Тамбовский государственный технический университет», профессор кафедры компьютерно-интегрированных систем в машиностроении

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования «Сибирский федеральный университет»

Защита состоится «23» января 2026 г. в 13:00 на заседании диссертационного совета 55.2.002.02 на базе ордена Трудового Красного Знамени Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ) по адресу: Москва, ул. Авиамоторная 8А.

С диссертацией можно ознакомиться в библиотеке и на сайте МТУСИ: https://dis.mtuci.ru/dis/02/

Автореферат разослан _____

Ученый секретарь
диссертационного совета 55.2.002.02
кандидат технических наук, доцент М.Г. Городничев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Современное цифровое пространство характеризуется стремительным ростом объемов мультимодального контента. По данным Brand Analytics в марте 2025 года число активных авторов в социальных медиа в России составило 83,8 млн, пользователи написали 1,93 млрд публичных сообщений – постов, репостов и комментариев. По сравнению с мартом 2024 года наблюдается значительный прирост активных пользователей на 22,5% и создаваемого ими контента на 4,3%.

Интернет играет ключевую роль в повседневной жизни, общество все больше зависит от цифровых технологий и интернет-пространства, что приводит к возникновению форм деструктивного поведения пользователей, и становится серьезной проблемой, наносящей вред не только отдельным пользователям, но и целым сообществам и организациям.

Современные технологии, такие как машинное обучение и искусственный интеллект, предоставляют новые возможности для анализа больших объемов данных.

Особенностью предлагаемого исследования является комплексный подход, объединяющий методы машинного обучения и искусственного интеллекта для анализа взаимосвязей между различными типами контента, что не только позволяет выявлять известные формы деструктивного поведения, но и обнаруживать новые, ранее не встречавшиеся паттерны, что особенно важно в условиях быстро меняющегося цифрового пространства.

В условиях усиления деструктивного информационного воздействия как извне, так и внутри страны – интеллектуальная обработка мультимодальных данных из сети Интернет приобретает особую значимость.

Степень разработанности исследования. Существующие работы отечественных исследователей внесли значительный вклад в развитие методов анализа деструктивного контента. Авторы В.В. Тельбух, А.В. Десятых, С.С. Андрушкевич, Л.П. Пилипенко предложили метод выявления деструктивного контента в интернет-ресурсах на основе комбинированного подхода с использованием Байесовского метода и поиска по словарю.

В.А. Минаев, А.Д. Реброва, А.В. Симонов рассматривали модели классификации текстового контента и методы предварительной обработки для выявления деструктивного поведения в социальных сетях. Авторы исследовали и применяли основные методы векторизации текстов: Bag of Words, TF-IDF, Word2vec.

Роскомнадзор использует автоматизированные системы для мониторинга интернета и выявления запрещенной законом информации, например система «Чистый интернет» состоит из различных модулей, анализирующих текстовую информацию, а система «Окулус» анализирует изображения и видео.

Зарубежные исследователи Z. Ping, Yu. Liu предположили, что мультимодальная интеллектуальная платформа способна эффективно анализировать данные, повышая экономическую выгоду.

Проведенный анализ существующих решений выявил ряд существенных ограничений. Предложенные методы ориентированы на анализ отдельных модальностей и не учитывают взаимосвязи между различными типами данных. Существующие системы преимущественно направлены на выявление уже

известных форм деструктивного контента и не обладают достаточной адаптивностью к новым, ранее не встречавшимся паттернам поведения.

Исследование выполнялось в рамках реализации ключевых государственных стратегических документов, направленных на развитие цифровых технологий и обеспечение безопасности, защиту граждан Российской Федерации. Работа соотносится с положениями Национальной стратегии «О развитии искусственного интеллекта на период до 2030 года» (утверждена Указом Президента РФ от 10 октября 2019 г. №490), Национальной стратегии «О комплексной безопасности детей в Российской Федерации на период до 2030 года» (утверждена Указом Президента РФ от 17 мая 2023 г. №358), Национальной стратегии безопасности (утверждена Указом Президента РФ от 2 июля 2021 г. №400) и относится к национальным интересам Российской Федерации: «сбережение народа России, развитие безопасного информационного пространства, защита российского общества от деструктивного информационно-психологического воздействия; укрепление традиционных духовно-нравственных ценностей».

Цель исследования – повышение эффективности и качества выявления деструктивного поведения пользователей в сети Интернет за счет разработки методов и средств интеллектуальной обработки мультимодальных данных.

Объект исследования – цифровой след пользователей в сети Интернет.

Предмет исследования – методы и средства обработки и классификации данных для выявления деструктивного поведения пользователей.

Для достижения поставленной цели были решены следующие **задачи исследования:**

1. Проведен анализ мультимодальных данных и деструктивного поведения в цифровой среде.
2. Разработан метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем.
3. Разработан кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента.
4. Разработана архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде.
5. Проведен анализ результатов эксперимента.

Методы исследования. В диссертации для решения указанных задач были использованы методы анализа и классификации данных (методы машинного обучения, байесовский анализ), для обработки мультимодальных данных (кросс-модальный анализ, методы предобработки данных), а также для проектирования систем (системный анализ, структурный анализ, объектно-ориентированное моделирование).

Научная новизна.

1. Метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем, *в отличие* от существующих подходов, автоматически формирует новый класс «Z» для потенциально неизвестных случаев деструктивного поведения и реализует интеллектуальную группировку данных по критериям угрозы, что *позволяет* сократить время выявления пользователей, состоящих в деструктивных сообществах в 125 раз, и расширить мультимодальность исследования цифрового следа за счет комплексного анализа данных из сети Интернет (2.3.8, п.7).

2. Кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента, *в отличие* от существующих методов, реализуется двухэтапная классификация контента за счет оценки вероятности проявления поведения пользователя, *позволяющий* определить принадлежность контента к деструктивному поведению и повысить точность выявления потенциально деструктивного пользователя цифровой платформы на 19% (2.3.8, п.4).

3. Архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде, *в отличие* от известных решений, единый аналитический контур включает сбор данных, специализированные искусственные нейронные сети для каждой модальности данных, кросс-модальную вероятностную классификацию и формирование рекомендаций, что *позволяет* повысить инвариантность к входным показателям без необходимой перенастройки моделей и сократить нагрузку на модераторов интеллектуальной системы (2.3.8, п.16).

Теоретическая и практическая значимость работы определяется установкой критериев интеллектуальной группировки данных по уровням угрозы с учетом мультимодальных признаков, установкой дополнительных признаков классификации контента и контекста и результатом интеграции разработанных методов и архитектуры является комплексное повышение эффективности и качества выявления деструктивного поведения пользователей в сети Интернет, что проявляется в оперативности, результативности, ресурсоэкономности, мультимодальности, точности и инвариантности.

Положения, выносимые на защиту:

1. Метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем.
2. Кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента.
3. Архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде.

Степень достоверности основных полученных результатов обеспечивается надежными исходными данными, адекватностью и корректностью применяемых методов исследования в области искусственного интеллекта и обработки данных. Достоверность подтверждается апробацией и внедрением полученных результатов в практику работы производственных и образовательных организаций.

Апробация работы. Основные положения и результаты диссертационного исследования докладывались и обсуждались на Российских и международных конференциях:

– 2024 4-ая Международная конференция «Technology Enhanced Learning in Higher Education (TELE 2024)» (г. Липецк);

– 2024 Всероссийская научно-практическая конференция «Актуальные проблемы деятельности подразделений УИС» (г. Воронеж);

– 2024 Международная научная конференция «Актуальные проблемы прикладной математики, информатики и механики» (г. Воронеж);

– 2025 Международная научно-практическая конференция «Техника и безопасность объектов УИС» (г. Воронеж).

Личный вклад автора. В ходе исследования автором лично были разработаны и реализованы ключевые подходы, направленные на решение поставленных научных задач. Разработан метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем [82]. Разработан кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента [73, 74]. Предложена архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде [72, 75, 83]. Проведен анализ мультимодальных данных и деструктивного поведения в цифровой среде.

Реализация и внедрение результатов исследования. Основные положения диссертационного исследования внедрены в практическую деятельность следующих организаций:

– ООО «НПК «КарбонГрупп» для анализа цифровых взаимодействий сотрудников;

– ООО «АЙТИ КЛАСС» для мониторинга корпоративных коммуникаций сотрудников;

– в учебный процесс Института информационных технологий РТУ МИРЭА.

Подтверждается соответствующими актами внедрения результатов диссертационного исследования, зафиксировавшими значительное повышение качества модерации контента и снижение конфликтных ситуаций.

Соответствие специальности. Тематика и результаты диссертации соответствуют следующим направлениям специальности: 2.3.8 – «Информатика и информационные процессы».

п.7 Разработка методов обработки, группировки и аннотирования информации, в том числе, извлеченной из сети интернет, для систем поддержки принятия решений, интеллектуального поиска, анализа.

п.4 Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видео контента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов.

п.16 Автоматизированные информационные системы, ресурсы и технологии по областям применения (научные, технические, экономические, образовательные, гуманитарные сферы деятельности), форматам обрабатываемой, хранимой информации. Системы принятия групповых решений, системы проектирования объектов и процессов, экспертные системы и др.

Публикации. По материалам диссертации опубликовано 17 научных работ (9 статей, 4 материала научных конференций, 4 свидетельства о регистрации программы для ЭВМ), 6 из которых опубликованы в научных журналах и изданиях, рекомендованных ВАК при Министерстве науки и высшего образования Российской Федерации, 1 − в сборнике индексируемом Scopus, в том числе 9 работ опубликованы без соавторов.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения, списка литературы и 2 приложения. Полный объем диссертации составляет 133 страницы, включая 48 рисунков и 36 таблиц. Список литературы содержит 131 наименование.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертации, определены цель, задачи, объект, предмет и методы исследования, показаны научная новизна, теоретическая и практическая значимость исследования.

В первой главе «Анализ мультимодальных данных и деструктивного поведения в цифровой среде» обосновывается фундаментальная роль мультимодальных данных и выявления деструктивного поведения пользователей.

Проведен анализ существующих унимодальных подходов выявления деструктивного поведения. Для текстовых данных применяются методы векторный (TF-IDF) и семантические (Word2Vec, BERT) для классификации тональности или токсичных высказываний. Для обработки фото и видео изображений используются сверточные нейронные сети (CNN) и детекторы объектов (YOLO), которые обеспечивают распознавание объектов в режиме реального времени. Анализ аудио включает распознавание речи (ASR) с последующим текстовым анализом. Основным недостатком унимодальных подходов является отсутствие анализа взаимосвязей между модальностями.

В связи с выявленными ограничениями исследован цифровой след пользователя как комплекс взаимосвязанных информационных потоков, включающий текстовые сообщения, аудиоматериалы, видеоролики и фотоизображения (Рисунок 1).

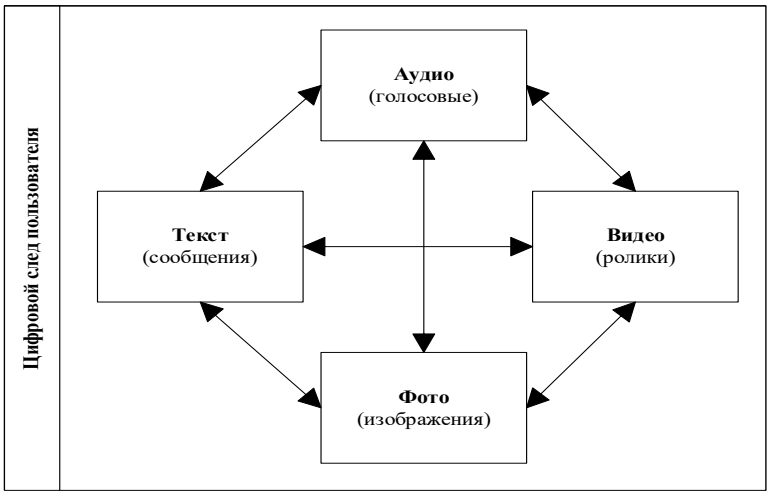

Рисунок 1 — Схема взаимосвязи цифровых следов пользователя

Мультимодальные данные представляют собой объединенную разнородную информацию, включающую текстовые сообщения, аудиозаписи, изображения и видеоролики о пользовательской активности.

Особенность данных заключается в их взаимодополняющем характере, анализ одного типа контента отображает неполную информацию поведения пользователя. В контексте выявления деструктивного поведения мультимодальность становится важной для выявления угроз в различных форматах.

Исследованы основные формы деструктивного поведения, а также их влияние на различные сферы, включая бизнес и образование (Таблица 1).

Таблица 1 – Влияние основных форм деструктивного поведения на примере организаций

Форма поведения	Бизнес-предприятия	Образовательные учреждения
Агрессивное поведение	Увеличение конфликтов, ухудшение качества принимаемых решений	Увеличение конфликтов, снижение качества образования, увеличение затрат на обеспечение безопасности
Кибербуллинг	Увеличение стрессовых ситуаций и депрессии, снижение продуктивности	Увеличение стрессовых ситуаций и депрессии, снижение продуктивности обучения
Распространение дезинформации	Репутационные риски компании, потеря доверия, снижение продаж	Потеря доверия, снижение качества образования

Обосновано, что деструктивное поведение в цифровой среде характеризуется различными причинами и факторами, которые можно разделить на несколько категорий (Таблица 2).

Таблица 2 – Деструктивное поведение в цифровой среде на примере факторов и причин возникновения

Фактор	Причина возникновения
Психологический	Низкая самооценка
	Эмоциональная нестабильность
Социальный	Влияние группы
	Социальная изоляция
Культурный	Нормы и ценности
	Медиа и контент
Технологический	Анонимность
	Скорость и доступность информации
Экономический	Финансовые мотивы
	Конкуренция

Применение технологий искусственного интеллекта позволяет обрабатывать большие объемы данных и выявлять скрытые закономерности в поведении пользователей.

Проведен обзор современных платформ передачи контента, таких как социальные сети, мессенджеры и форумы.

Установлено, что каждая платформа имеет уникальные характеристики, которые необходимо учитывать при разработке аналитических инструментов (Таблица 3).

Таблица 3 – Сравнительный анализ цифровых платформ по параметрам контента

Параметр	Социальные сети	Мессенджеры	Форумы
Преобладающий тип данных	Текст, изображения, видео	Текст, аудио, изображения, видео	Текст
Уровень модерации	Средний	Низкий	Средний
Скорость распространения	Высокая	Высокая	Низкая

Рассмотрены ключевые аспекты мультимодальности, включая особенности семантического взаимодействия текстовых, аудио, фото и видеоданных (Таблица 4).

Таблица 4 – Мультимодальные данные на примере семантического взаимодействия

Текст	**Аудио/Видео сопровождение**	**Итоговая интерпретация**
«Отличная работа»	Саркастическая интонация	Критика, недовольство
«Это интересно»	Отсутствие эмоций в голосе, скучающее выражение лица	Скрытое безразличие
Нейтральное описание	Агрессивные жесты на видео	Скрытая угроза

Показано, что мультимодальный подход позволяет более точно интерпретировать поведение пользователей, особенно в контексте скрытых угроз.

Основные проблемы, связанные с анализом мультимодальных данных, включают контекстную зависимость, ограниченный доступ к данным в закрытых коммуникациях и сложность обработки разнородных форматов, что требует дальнейшего изучения и разработки решений, таких как метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем и кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента.

Таким образом, установлено, что для разработки методов обработки научно-методических и системно-технических средств интеллектуальной системы необходимы анализ и цифровая обработка мультимодальных данных, характеризующих деструктивное поведение пользователей в сети Интернет, что определяет научную задачу исследования.

Во <u>второй главе</u> «Разработка метода обработки мультимодальных данных из сети Интернет для интеллектуальных систем» представлен разработанный метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем.

К этапам обработки данных метода относятся: 1) сбор и агрегация данных из различных источников; 2) параллельная предобработка и анализ каждой модальности с помощью специализированных моделей; 3) интеллектуальная группировка полученных признаков по критериям угрозы (К1-К3); 4) оценка риска для формирования потенциального класса «Z»; 5) представление результатов.

В рамках данного решения мультимодальные данные объединяют: текстовые данные, медиа данные и социальную активность пользователей (Рисунок 2).

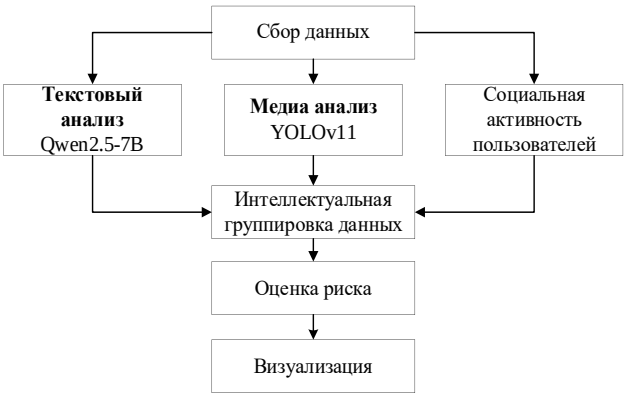

Рисунок 2 — Этапы обработки мультимодальных данных

Обработка данных объединяет три модуля, результаты которых группируются по критериям.

Логическая диаграмма (Рисунок 3) демонстрирует объединение критериев для комплексной оценки риска на этапе интеллектуальной группировки.

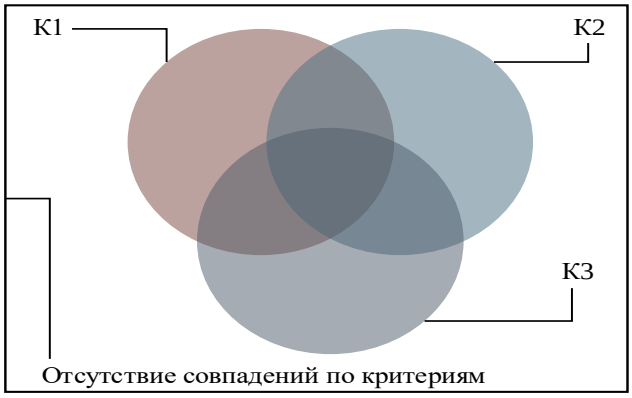

Рисунок 3 — Логическая диаграмма критериев интеллектуальной группировки

Критерий 1 (К1) – содержит текстовые данные пользователя и возможные отклонения в них. Критерий 2 (К2) – содержит медиа данные пользователя и возможные отклонения в них. Критерий 3 (К3) – содержит деструктивные сообщества пользователя и возможные участия в них.

Аналитическое описание метода к процессу присвоения категорий по пользователю на основе анализа его мультимодальных данных по критериям К1, К2 и К3, представлена в формулах 1-8.

Шаг 1. Векторизация и оценка по критериям. Пусть u – конкретный пользователь. Его цифровой след представлен тремя типами данных:

$D_{text}(u)$ – текстовые данные (посты, комментарии).

$D_{media}(u)$ – медиаданные (изображения, видео).

$D_{social}(u)$ – данные о социальных связях (членство в сообществах).

Для каждого критерия существует специализированная модель, которая преобразует сырые данные в вероятность наличия деструктивных признаков.

1. Критерий К1 (Текстовые данные):

$$f_{К1}: D_{text}(u) \to p_1(u), \#(1)$$

где $p_1(u) \in [0,1]$ – вероятность того, что текстовый контент пользователя содержит деструктивные элементы.

2. Критерий К2 (Медиаданные):

$$f_{К2}: D_{media}(u) \to p_2(u), \#(2)$$

где $p_2(u) \in [0,1]$ – вероятность того, что медиаконтент пользователя содержит деструктивные элементы.

3. Критерий К3 (Социальная активность пользователей):

$$f_{К3}: D_{social}(u) \to p_3(u), \#(3)$$

где $p_3(u) \in [0,1]$ – вероятность того, что пользователь состоит в деструктивных сообществах (нормализованное количество деструктивных сообществ).

В результате для пользователя u формируется вектор риска:

$$p(u) = \big(p_1(u), p_2(u), p_3(u)\big)\#(4)$$

Шаг 2. Бинаризация оценок по пороговому значению. Для принятия решения о «совпадении» критерия вводится порог r, что позволяет преобразовывать непрерывные вероятности в бинарные решения.

Для каждого критерия $k \in \{1,2,3\}$ определяется бинарная переменная:

$$ck(u) = \begin{cases} 1, \text{если } pk(u) \geq r \text{ (критерий совпал, риск обнаружен)} \\ 0, \text{если } pk(u) < r \text{ (критерий не совпал, риск не обнаружен)} \end{cases} \quad (5)$$

Таким образом, пользователь u описывается бинарным вектором совпадений:

$$\mathrm{c}(u) = \left(\mathrm{c}_1(u), \mathrm{c}(u), \mathrm{c}_3(u)\right)\#(6)$$

Шаг 3. Правило интеллектуальной группировки (присвоение категории риска). Категория риска *R(u)* присваивается пользователю u на основе нормы (длины) его бинарного вектора совпадений *c(u)*. Норма $L1$ вектора (вес Хемминга для бинарных векторов) равна количеству единиц в нем:

$$||\mathrm{c}(u)||_1 = \mathrm{c}_1(u) + \mathrm{c}(u) + \mathrm{c}_3(u)\#(7)$$

Правило группировки, представленное в таблице 5:

$$R(u) = \begin{cases} \text{Критический, если } ||\mathrm{c}(u)||_1 = 3 \\ \text{Высокий, если } ||\mathrm{c}(u)||_1 = 2 \\ \text{Средний, если } ||\mathrm{c}(u)||_1 = 1 \\ \text{Низкий, если } ||\mathrm{c}(u)||_1 = 0 \end{cases} \#(8)$$

Приведены категории риска и их критерии на основе интеллектуальной группировки (Таблица 5).

Таблица 5 – Критерии интеллектуальной группировки для оценки риска

Категория риска	**Критерии**
Критический	Совпадение по всем 3 критериям
Высокий	Совпадение по 2 любым критериям из 3
Средний	Совпадение по 1 критерию из 3
Низкий	Отсутствие совпадений по критериям

Предложенный метод выявляет новые формы деструктивного контента через мультимодальный анализ и интеллектуальную группировку. Класс «*Z*» — новая форма деструктивного поведения или не классифицируемые действия пользователей.

Метод позволяет выявить противоречия между модальностями или не классифицируемыми действиями пользователей, далее данные группируются по схожим признакам. Новый кластер передается в блок «Экспертная верификация» для подтверждения его деструктивной направленности. На данном этапе новый класс «Z» добавляется в классификатор или исключается из анализа.

Схема на рисунке 4 демонстрирует процесс выявления нового класса.

Рисунок 4 — Схема выявления нового класса «*Z*»

Для подтверждения эффективности разработанного метода обработки мультимодальных данных было проведено тестирование. В качестве объекта сравнения выбрана действующая система анализа цифровых взаимодействий сотрудников в ООО «НПК «КарбонГрупп».

Под «сообществом» в данном контексте понимается сообщество в социальной сети Интернет, в котором состоит пользователь. Общее количество пользователей составило 2000. Для анализа деструктивной активности рассматривались их подписки и участие в сообществах. Каждый пользователь в среднем состоит в 25-26 сообществах, общее количество для проверки составило 51000 сообществ. (Таблица 6).

Таблица 6 – Результаты сравнения существующего подхода модерации и предложенного метода

Критерий	**Существующий подход**	**Предложенный метод**	**Эффективность**
Объем данных (2000 пользователей)	Выборочная проверка (50-100 пользователей)	Все пользователи	Полный охват пользователей
Общее время обработки (51000 сообществ)	142 ч	1,14 ч	В 125 раз
Скорость обработки (1 сообщество)	10 сек	0,08 сек	
Результат для эксперта	Список групп для ручной проверки	Готовый отчет: пользователь и количество деструктивных групп	Структурированный результат для эксперта

Полученные результаты, представленные во второй главе, демонстрируют высокую эффективность применения предложенного метода в различных областях, включая мониторинг социальных сетей, обеспечение безопасности и модерацию контента.

Для достижения точности классификации деструктивного поведения в третьей главе представлен кросс-модальный вероятностный метод классификации.

В <u>третьей главе</u> «Кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента» обосновывается метод, сочетающий цифровую обработку контента и его контекста, с последующим вычислением условной вероятности принадлежности контент-объекта пользователя к его форме поведения.

Разрабатываемый подход основан на двухэтапном анализе для качественного выявления деструктивного поведения пользователей. На этапе анализа признаков выполняется кросс-модальный анализ контента и контекста, где учитываются не только сами данные, но и их взаимосвязи. Кросс-модальный подход к обработке данных, отличается от существующих унимодальных методов способностью анализировать и выявлять взаимосвязи между различными типами контента.

На этапе расчета вероятности метода система переходит к объектно-вероятностной классификации, которая формализует принятие решений о деструктивном поведении через взаимосвязь мультимодальных признаков. На рисунке 5 продемонстрированы этапы разработанного метода.

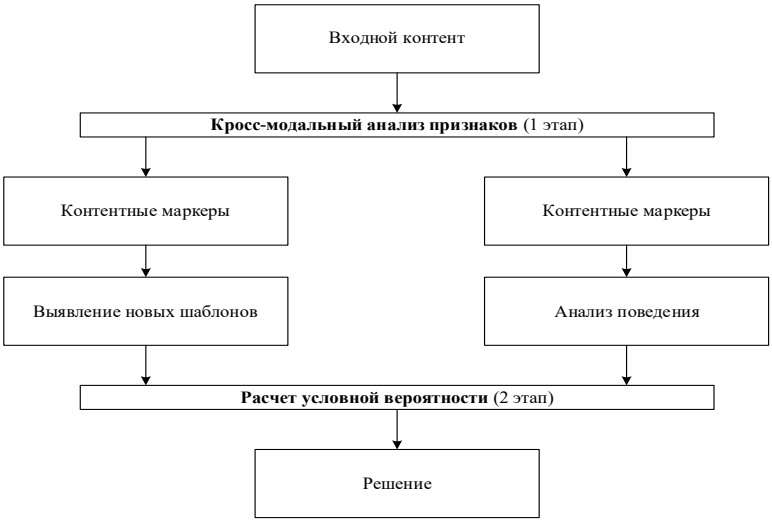

Рисунок 5 – Схема этапов кросс-модального вероятностного метода классификации

В рамках данного исследования критерий отнесения к деструктивному поведению является не внешним правилом, а внутренним свойством обучающей выборки. Граница между деструктивным и не деструктивным контентом проводится: на этапе обучения через разметку данных или на этапе классификации по сходству эталонных примеров из обучающей выборки. Вектор признаков X формируется как совокупность контентных и контекстных признаков, фрагменты которых для различных модальностей приведены в таблице 7.

Таблица 7 – Фрагменты контентных и контекстных признаков

Тип контента	**Контентные признаки**	**Контекстные признаки**
Текст	Токсичная лексика, угрозы, угрозы, деструктивные призывы и лозунги	Аномальная активность в ночное время, высокая частота сообщений, координированные действия пользователей
Аудио	Крики, ненормативная лексика, агрессивная интонация, угрозы	Фоновые звуки (выстрелы, гипноз), неестественные паузы или модификации голоса
Видео	Демонстрация насилия, опасные челленджи, запрещенная символика, противоправные действия, дипфейки	Аномальная длительность роликов, повторяющиеся паттерны в разных видео от одного автора, признаки монтажа или глубокой обработки
Изображения	Символика экстремизма, запрещенная атрибутика	Признаки отредактированного фото, неестественные цветовые оттенки или следы обработки

Каждый контент-объект рассматривается как динамическая система признаков. Формула (9) для расчета вероятности деструктивности:

$$P(Z|X) = \frac{P(X|Z) \times P(Z)}{P(X)}, \#(9)$$

где, P(Z|X) – вероятность, что контент относится к классу деструктивности Z при данных признаков X, P(X|Z) – вероятность наблюдения признаков X для класса

Z, P(Z) – базовая распространенность класса Z, P(X) – общая вероятность признаков X.

В таблице 8 представлены компоненты формулы вероятностной классификации.

Таблица 8 – Компоненты формулы вероятностной классификации

Обозначение	**Название**	**Описание**	**Как вычисляется**
P(Z|X)	Апостериорная вероятность	Вероятность, что контент относится к классу Z при наблюдаемых признаках X	Рассчитывается по формуле Байеса
P(X|Z)	Правдоподобие	Вероятность наблюдения признаков X для класса Z	Статистика по обучающей выборке
P(Z)	Априорная вероятность	Базовая распространенность классов в данных $P(Z = k) = \frac{N_k}{N}, k \in \{\text{дестр}, \text{не дестр}\}$	Доля объектов каждого класса в обучающем наборе
P(X)	Нормирующая константа	Общая вероятность наблюдения признаков X во всех классах	Сумма P(X|Z)*P(Z) по всем классам

Формула (9) представляет собой байесовский классификатор, который вычисляет вероятность того, что анализируемый контент является деструктивным по классу Z, учитывая все его признаки X.

Таким образом, окончательное решение о принадлежности контента к деструктивному классу является результатом интерпретации апостериорной вероятности.

На рисунке 6 продемонстрирована схема кросс-модального вероятностного метода классификации.

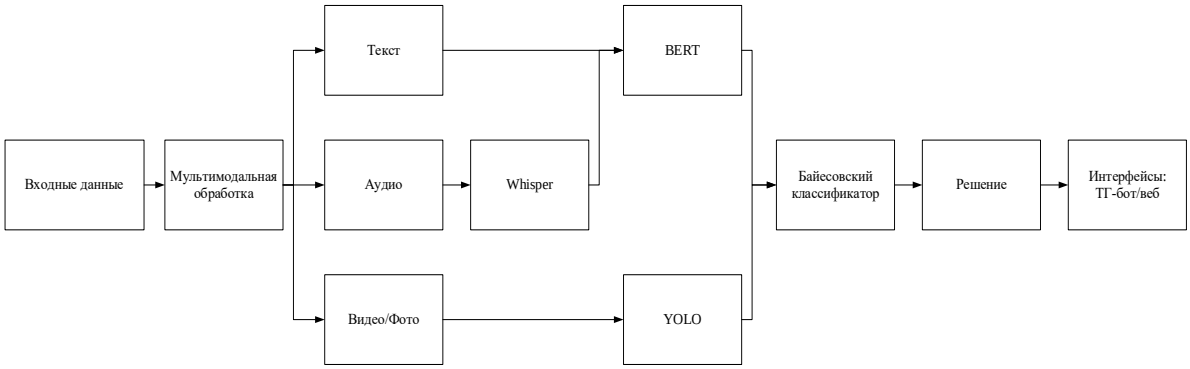

Рисунок 6 – Схема кросс-модального вероятностного метода классификации

На рисунке 7 продемонстрирован результат работы кросс-модального вероятностного метода классификации.

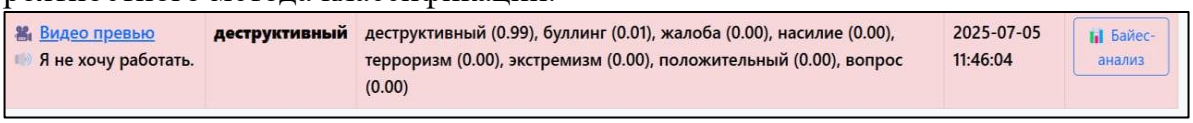

Рисунок 7 – Результат явного обнаружения деструктивного поведения пользователя

Разработанный метод с вероятностью 99% идентифицировал данное сообщение как проявление деструктивного поведения. Несмотря на отсутствие

явных маркеров агрессии или нарушений, корректно была распознана скрытая форма деструктивного поведения – саботаж трудовой деятельности или демотивация.

Для подтверждения качества кросс-модального вероятностного метода классификации было проведено тестирование в корпоративном мессенджере. В качестве объекта сравнения выбрана действующая система выявления деструктивного поведения в корпоративном мессенджере ООО «АЙТИ КЛАСС».

Расчет показателей проводился на наборе данных объемом 1257 сообщений с экспертной разметкой (Таблица 9).

Таблица 9 – Объем набора данных по типу сообщения из корпоративного мессенджера ООО «АЙТИ КЛАСС»

Тип сообщения	**Текстовые сообщения**	**Видео сообщения**	**Фото изображения**	**Аудио сообщения**	**Всего сообщений**
Объем	857	117	194	89	1257

Выражение метрики «точность»:

Для существующего подхода: $\text{Accuracy} = \frac{527+353}{527+353+215+162} = \frac{880}{1257} = 0{,}70$

Для предложенного метода: $\text{Accuracy} = \frac{765+354}{765+354+68+70} = \frac{1119}{1257} = 0{,}89$

Выражение метрики (F1-score):

Для предложенного метода: $\text{Precision} = \frac{765}{765+68} = 0{,}918$

Для предложенного метода: $\text{Recall} = \frac{765}{765+70} = 0{,}916$

Для предложенного метода: $\text{F1} - \text{score} = 2 \times \frac{0{,}918 \times 0{,}916}{0{,}916 + 0{,}918} = 0{,}917$

Существующий подход подразумевал оценку классификации по метрике Recall, численное значение которой составляло 0,60. Для такого значения Recall теоретическое максимальное значение метрики F1-score составляет 0,75 (Таблица 10).

Результаты демонстрируют, что предложенный метод обеспечивает значительное улучшение по всем ключевым метрикам за счет комплексного анализа мультимодальных данных.

Таблица 10 – Результаты сравнения существующего подхода в ООО «АЙТИ КЛАСС» и предложенного метода

Критерий	**Существующий подход в ООО «АЙТИ КЛАСС»**	**Предложенный метод**	**Результат**
Точность	70%	89%	+19%
Класс «Z»	60% (*Recall*) (<=0,75 *F1-score*)	92% (*F1-score*)	+>17%

Вышеописанный кросс-модальный вероятностный метод классификации цифровой обработки контента встраивается в интеллектуальную информационную систему. Для визуализации результатов используются отчеты и рекомендации, обеспечивающие принятие обоснованных решений.

В <u>четвертой главе</u> «Архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде» данная

архитектура анализирует поведение пользователей в цифровой среде, обрабатывая мультимодальные данные.

Определяющей характеристикой предлагаемой архитектуры выступает способность к обработке данных различной модальности, что принципиально важно для работы с гетерогенным цифровым контентом (текст, речь, видео, изображения).

Критерии архитектурного выбора предполагают: интеграцию с социально-коммуникационными платформами, оптимальность под ограниченные вычислительные ресурсы, поддержку детектирования известных и новых сценариев деструктивного поведения (класс «Z»), соблюдение законодательства в сфере обработки данных.

Проектируемая интеллектуальная информационная система организована на базе архитектуры сквозного конвейера, в котором контур сбора исходных данных из многоканальной цифровой среды переходит в модуль мультимодальной обработки с интеграцией представлений, после чего выполняется кросс-модальная вероятностная классификация в унифицированном пространстве признаков, а завершается процесс подсистемой формирования рекомендаций, ориентированной на прикладные цели пользователя. На рисунке 8 представлена диаграмма последовательности, демонстрирующая логику выработки рекомендаций системой мультимодальной обработки в цифровой среде и прослеживающая взаимодействие компонент от поступления данных до выдачи результата.

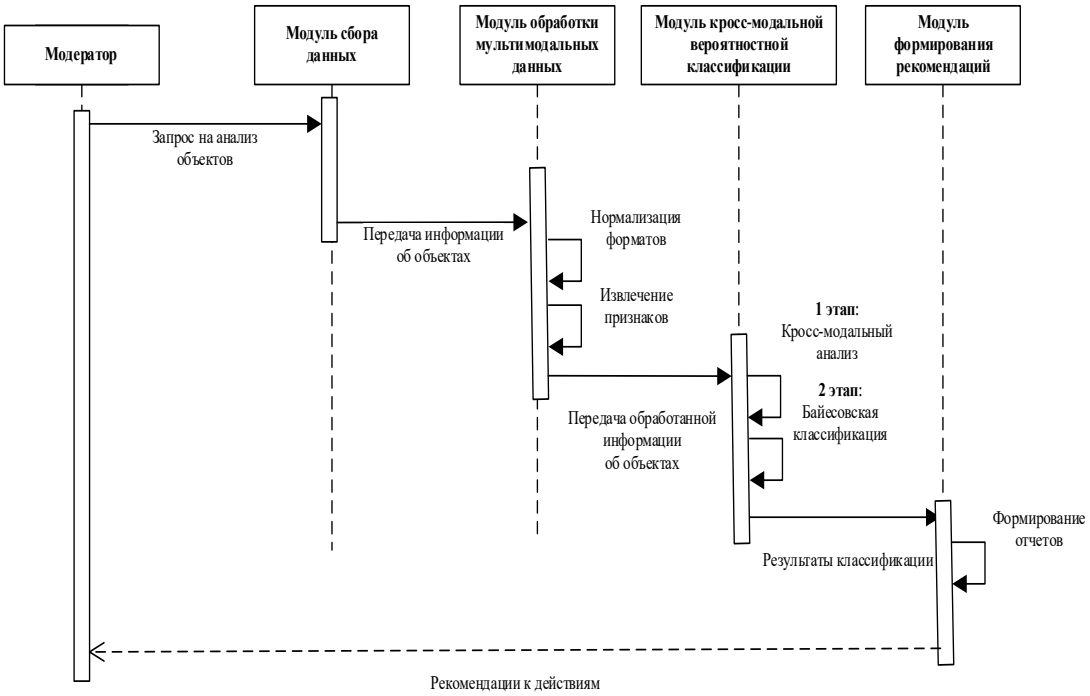

Рисунок 8 – Диаграмма последовательности формирования рекомендаций в интеллектуальной информационной системе

Диаграмма последовательности визуализирует согласованное взаимодействие компонентов системы, отслеживая маршрут от момента инициирования запроса модератором до этапа вычисления и выдачи рекомендаций и фиксируя причинно-временные зависимости между операциями. Архитектура интеллектуальной информационной системы, реализующей обработку

мультимодальных данных в цифровой среде, представлена на рисунке 9 и отражает целостную организацию функциональных контуров и их интеграционные связи, обеспечивающие непрерывность конвейера обработки и устойчивость результатов.

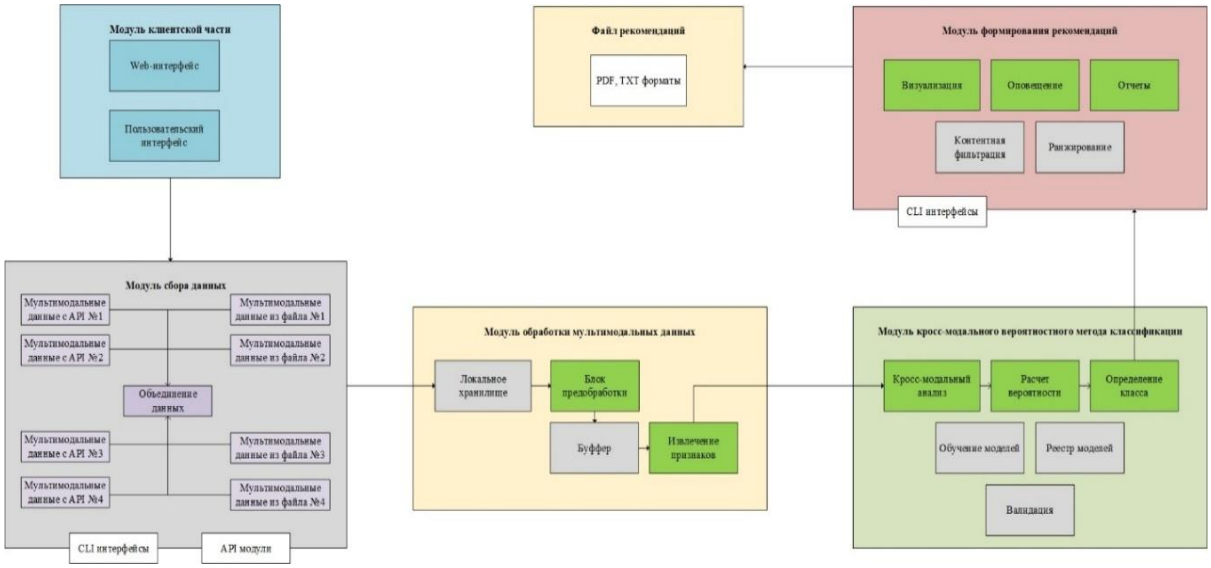

Рисунок 9 – Архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде

Схема архитектуры интеллектуальной информационной системы демонстрирует действие принципа однонаправленного прохождения данных, при котором поток инициируется подсистемами сбора, последовательно трансформируется на стадиях обработки и завершается консистентным представлением результатов для потребителей. Каждый компонент реализует строго определённую функциональность и взаимодействует с другими компонентами через формализованные интерфейсы, что обеспечивает модульность, слабую связанность и технологичность сопровождения. На рисунке 10 отображён единый аналитический контур архитектуры, который связывает ключевые этапы преобразования данных и обеспечивает сквозную трассировку и воспроизводимость вычислений на всём протяжении конвейера.

Рисунок 10 – Единый аналитический контур архитектуры интеллектуальной информационной системы

Единый аналитический контур обеспечивает сквозную согласованность обработки информации и формирует устойчивость к межмодальной гетерогенности входных данных, одновременно поддерживая адаптацию к ранее не наблюдавшимся формам угроз в полностью автоматическом режиме без привлечения ручных корректировок. На рисунке 11 представлена концептуальная схема процесса обработки мультимодальных данных в архитектуре интеллектуальной информационной системы, которая визуализирует путь преобразования и интеграции сигналов от момента поступления до получения интерпретируемых результатов.

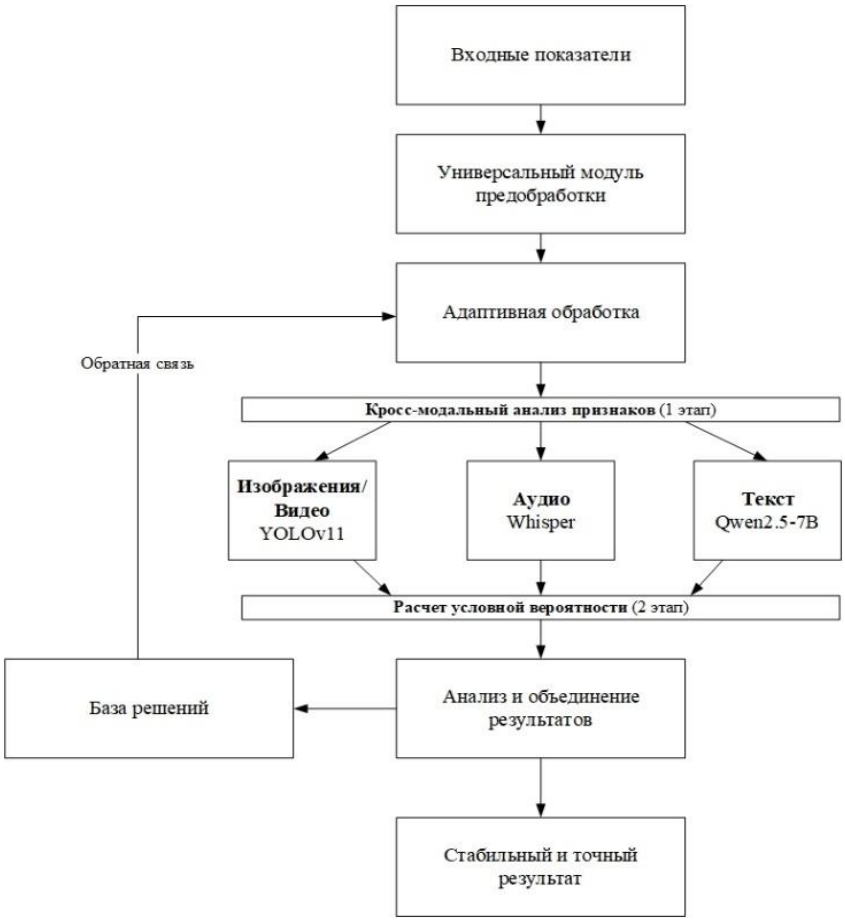

Рисунок 11 – Схема процесса обработки мультимодальных данных в архитектуре интеллектуальной информационной системы

Представленная схема демонстрирует, как архитектурная организация системы обеспечивает повышение инвариантности к вариативности входных показателей за счёт согласованного комплекса функциональных решений. Универсальный модуль предобработки реализует первичную нормализацию и приведение разнородных источников к согласованным формальным представлениям, что задаёт целостность трактовки на следующих этапах обработки. Специализированные обработчики по типам модальностей, построенные на адаптивных моделях искусственного интеллекта, проявляют робастность к колебаниям входных признаков благодаря предобучению на репрезентативных корпусах и последующей калибровке в целевых доменах. В качестве входных показателей рассматриваются мультимодальные данные,

поступающие из различных цифровых платформ, включая мессенджеры и социальные сети, где неоднородность форматов и условий наблюдения компенсируется унифицированным контуром подготовки и межмодальной интеграцией.

База решений формирует устойчивый слой памяти системы, где аккумулируются результаты экспертной валидации, контекст применения и характеристики неопределенности модели. Такое хранилище выступает опорной точкой для согласования представлений данных и для эволюции поведенческих правил, поскольку каждое зафиксированное решение связывается с источником, модальностью, версией модели и операционными параметрами, которые влияют на итоговую интерпретацию. Наличие структурированного корпуса подтвержденных примеров создает основу для стабильной генерализации при колебаниях входных распределений и для корректного переноса знаний в новые домены.

В таблице 11 представлены результаты применения архитектуры интеллектуальной информационной системы, подтверждающие инвариантность и сокращение нагрузки на модераторов.

Таблица 11 – Результаты сравнения существующих подходов и предложенного решения

Задача	**Существующие подходы**	**Предложенное решение**
Обработка разноформатных данных	Разрозненные системы для каждого типа данных, отсутствие единого стандарта обработки	Универсальный модуль предобработки со специализированными искусственными нейронными сетями для каждой модальности
Выявление новых угроз (класс «Z»)	Статические модели с возможностью ручного обновления при появлении угроз	Кросс-модальный вероятностный метод классификации и механизм обратной связи для непрерывного обучения
Энергоэффективность и производительность	Ресурсоемкие алгоритмы, не оптимизированные для работы в условиях ограниченных вычислительных ресурсов	Использование ресурсоэффективных моделей искусственного интеллекта
Инвариантность к входным данным	Обработка конкретных форматов данных, необходимость перенастройки под новые платформы	Адаптивная интеллектуальная обработка разнородных данных

На рисунке 12 продемонстрировано сравнение нагрузки на модератора при существующем процессе модерации и с применением разработанной архитектуры интеллектуальной информационной системы.

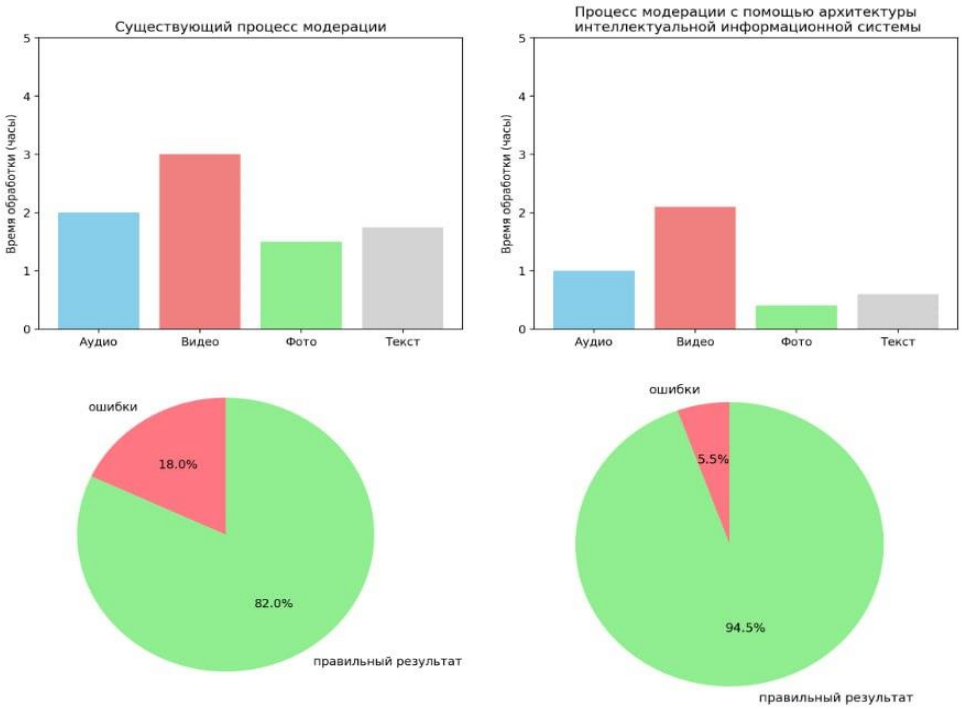

Рисунок 12 – Сравнение нагрузки на модератора

Действующий процесс модерации характеризуется высокой нагрузкой на модератора из-за преобладания ручной проверки данных. Разработанная архитектура интеллектуальной информационной системы автоматизирует обработку 94,5% общего потока, сокращает время прохождения заявок и освобождает модератора от рутинных операций, что повышает результативность с 82% до 94,5%.

Процесс модерации в рамках архитектуры интеллектуальной информационной системы автоматизирует основную часть обработки контента, оставляя модераторам лишь наиболее сложные и спорные случаи, доля которых составляет 5,5%. Тем самым архитектура обеспечивает интерпретируемость выводов для модераторов и адаптивность к разнородным типам данных цифровой среды, включая текст, аудио, видео и изображения.

Далее представлен сводный анализ результатов эксперимента (Таблица 12).

Таблица 12 – Сводный анализ результатов эксперимента

Характеристика	**Результат**
Оперативность	Сокращение времени выявления в 125 раз (51000 сообществ и 2000 пользователей
Результативность	Определение деструктивного контента 99%, включая скрытые формы
Ресурсоэкономность	Сокращение трудозатрат модератора за счет автоматизации с 82% до 5,5%
Мультимодальность	Охват всех типов контента и их взаимосвязи (текст, фото, аудио, видео)
Точность	Рост точности на 19% (1257 сообщений в корпоративном мессенджере)
Инвариантность	Устойчивость к изменениям в цифровой среде за счет механизма обратной связи, обработка данных из

	разных платформ (социальные сети, корпоративные мессенджеры)

Повышение эффективности и качества выявления деструктивного поведения в сети Интернет подтверждено анализом экспериментальных результатов и достигается применением комплекса решений, в ядре которого кросс-модальный вероятностный метод классификации аудио-, текстовых, фото- и видеоданных, метод обработки мультимодальных интернет-данных для интеллектуальных систем и архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде.

В **заключении** подведены итоги диссертационного исследования, сформулированы **основные результаты и выводы**.

В диссертации решена научная задача, состоящая в разработке методов обработки научно-методических и системно-технических средств интеллектуальной системы на основе анализа и цифровой обработки мультимодальных данных, характеризующих деструктивное поведение пользователей в сети Интернет.

Основные результаты проведенного исследования заключаются в следующем:

1. Проведен анализ мультимодальных данных и деструктивного поведения в цифровой среде. Показано, что мультимодальный подход позволяет более точно интерпретировать поведение пользователей, особенно в контексте скрытых угроз.

2. Разработан метод обработки мультимодальных данных из сети Интернет для интеллектуальных систем. В отличие от существующих подходов, метод автоматически формирует новый класс «Z» для потенциально неизвестных (ранее не классифицированных) случаев деструктивного поведения. Разработанный метод сокращает время выявления пользователей, состоящих в деструктивных сообществах в 125 раз, и расширяет мультимодальность исследования цифрового следа за счет комплексного анализа данных из сети Интернет. Практическая значимость метода подтверждена результатами экспериментальных исследований, проведенных на базе РТУ МИРЭА и ООО «НПК «КарбонГрупп» (акты о внедрении результатов диссертационного исследования в приложении диссертации).

3. Разработан кросс-модальный вероятностный метод классификации цифровой обработки аудио, текстовых данных, фото и видеоизображений контента. В отличие от существующих методов, реализуется двухэтапная классификация контента за счет оценки вероятности проявления поведения пользователя, позволяющий определить принадлежность контента к деструктивному и повысить точность выявления потенциально деструктивного пользователя цифровой платформы на 19%. Практическая значимость подтверждена результатами внедрения в ООО «АЙТИ КЛАСС» (акт о внедрении результатов диссертационного исследования в приложении диссертации).

4. Разработана архитектура интеллектуальной информационной системы обработки мультимодальных данных в цифровой среде. В отличие от известных решений, единый аналитический контур включает сбор данных, специализированные искусственные нейронные сети для каждой модальности данных, кросс-модальную вероятностную классификацию и формирование рекомендаций, что позволяет повысить инвариантность к входным показателям без необходимой перенастройки моделей и сократить нагрузку на модераторов

интеллектуальной системы. (свидетельство о регистрации программы № 2025668422 от 15.07.2025 «Интеллектуальная информационная система обработки мультимодальных данных» и свидетельство о регистрации программы № 2025681455 от 14.08.2025 «Интеллектуальная информационная система обработки данных в образовательной сфере»).

5. Анализ результатов эксперимента всесторонне доказывает, что цель диссертационного исследования достигнута, а поставленные задачи решены. Разработанные методы и архитектура обеспечивают комплексное повышение эффективности и качества выявления деструктивного поведения в сети Интернет, что проявляется в оперативности, результативности, ресурсоэкономности, мультимодальности, точности и инвариантности.

Проведенное исследование вносит значительный вклад в развитие теории и практики интеллектуальной обработки мультимодальных данных из сети Интернет.

Дальнейшие исследования могут быть направлены на масштабирование интеллектуальной системы в рамках предиктивного выявления угроз в коммерческих и образовательных организациях для обеспечения корпоративной безопасности в сети Интернет.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК при Минобрнауки России

1. Тетерин Н.Н. Влияние новой информационной системы на управление организационными системами / Н.Н. Тетерин, Ю.В. Чебернова // Конкурентоспособность в глобальном мире: экономика, наука, технологии. – 2023. – № 6. – С. 71-72. **(Перечень рецензируемых научных изданий № 1563 от 01.07.2025).**

2. Тетерин Н.Н. К вопросу выявления деструктивного поведения на основе анализа текстовых данных с использованием методов машинного обучения / Н.Н. Тетерин, В.В. Смоленцева // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. – 2025. – № 4/2. – С. 126-129. **(Перечень рецензируемых научных изданий № 2580 от 01.07.2025).**

3. Тетерин Н.Н. Концепция архитектуры информационной системы выявления деструктивного поведения в организационных системах // System Analysis & Mathematical Modeling. – 2025. – Т. 7, № 2. – С. 204–213. **(Перечень рецензируемых научных изданий № 151 от 01.07.2025).**

4. Тетерин Н.Н. / Разработка архитектуры интеллектуальной информационной системы обработки мультимодальных данных в цифровой образовательной среде // Экономика и качество систем связи. – 2025. – № 3(37). – С. 158-169. **(Перечень рецензируемых научных изданий № 1730 от 01.07.2025).**

5. Тетерин Н.Н. Методика обработки и интеллектуальной группировки мультимодальных данных // Научно-аналитический журнал

«Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». – 2025. – №3. – С. 189-195. **(Перечень рецензируемых научных изданий № 1879 от 01.07.2025).**

6. Тетерин Н.Н. Обработка и аннотирование данных в распределенной системе интеллектуального анализа видеопотоков для детекции деструктивного поведения. / Н.Н. Тетерин, Т.Е. Смоленцева // Computational Nanotechnology. – 2025. – Т. 12. – №3. – С. 178-183. **(Перечень рецензируемых научных изданий № 27 от 01.07.2025).**

Публикации в изданиях, входящих в международные наукометрические базы Web of Science и Scopus

7. N. N. Teterin, N. Y. Borzykh, T. E. Smolentseva and A. V. Kalach, "Algorithm for Optimal Selection of a Strategy for Designing Corporate Information Systems Using Multi-criteria Optimization," 2024 4th International Conference on Technology Enhanced Learning in Higher Education (TELE), Lipetsk, Russian Federation, 2024, pp. 106-110.

Патенты и свидетельства о регистрации программных продуктов

8. Свидетельство о государственной регистрации программы для ЭВМ № 2025668422 Российская Федерация. «Интеллектуальная информационная система обработки мультимодальных данных»: заявл. 04.07.2025: опубл. 15.07.2025 / Н. Н. Тетерин.

9. Свидетельство о государственной регистрации программы для ЭВМ 2025681455 Российская Федерация. «Интеллектуальная информационная система обработки данных в образовательной сфере»: заявл. 04.07.2025: опубл. 14.08.2025 / Н. Н. Тетерин.

10. Свидетельство о государственной регистрации программы для ЭВМ № 2025683295 Российская Федерация. Программная реализация классификации текстовых данных: заявл. 20.08.2025: опубл. 02.09.2025 / Н.Н. Тетерин.

11. Свидетельство о государственной регистрации программы для ЭВМ № 2025687164 Российская Федерация. Программная реализация метода двухэтапной классификации контента: заявл. 24.09.2025: опубл. 08.10.2025 / Н.Н. Тетерин.

Публикации в иных изданиях

12. Тетерин Н.Н. К вопросу формализации задачи выявления деструктивного поведения с применением технологий искусственного интеллекта / Н.Н. Тетерин, В.В. Смоленцева // Тенденции развития науки и образования. – 2024. – № 114-10. – С. 81-83.

13. Тетерин Н.Н. Общие вопросы анализа деструктивного поведения пользователей в социальных сетях // Актуальные проблемы деятельности подразделений уголовно-исполнительной системы: Сборник материалов

Всероссийской научно-практической конференции. В 3-х томах, Воронеж, 24 октября 2024 года. – Воронеж: ООО Издательско-полиграфический центр «Научная книга», 2024. – С. 96-99.

14. Тетерин Н.Н. Концептуальный подход классификации деструктивного поведения с применением технологий искусственного интеллекта / Н.Н. Тетерин, В.В. Смоленцева // Актуальные проблемы прикладной математики, информатики и механики: Сборник трудов Международной научной конференции, Воронеж, 02–04 декабря 2024 года. – Воронеж: Научно-исследовательские публикации, 2025. – С. 326-329.

15. Тетерин Н.Н. Комплексный подход к разработке методических и алгоритмических средств поддержки принятия решений в управлении деструктивным поведением интернет-пользователей // Техника и безопасность объектов уголовно-исполнительной системы: Сборник материалов Международной научно-практической конференции. В 4-х томах, Воронеж, 14–15 мая 2025 года. – Воронеж: ИП Копыльцов П.И., 2025. – С. 573-575.

16. Тетерин Н.Н. Обработка данных из сети Интернет для интеллектуальных систем / Н.Н. Тетерин, Е.В. Климкин // Современные достижения технических наук: новые горизонты знаний, нанотехнологии: Сборник научных статей – 2025. – С. 35-38.

17. Тетерин Н.Н. Кросс-модальный вероятностный метод классификации цифровой обработки мультимодальных данных / Н.Н. Тетерин, Т.Е. Саратова // Безопасность. Управление. Искусственный интеллект. – 2025. – № 4. – С. 9-13.