

Разработка антифейк-сервиса образовательного назначения для анализа русскоязычных новостных текстов

Цель исследования. Актуальность данного исследования обусловлена устойчивым ростом дезинформации в цифровой среде, что делает разработку систем автоматизированной верификации контента не только критически значимым инструментом обеспечения национальной информационной безопасности, но и важной цифровой технологией образовательного назначения для повышения цифровой грамотности современных пользователей. Цель работы заключается в разработке и обосновании архитектуры антифейк-сервиса как специализированного цифрового инструмента образовательного назначения. В рамках исследования решаются следующие задачи: сравнительная оценка эффективности нейросетевых моделей для анализа новостей на русском языке и формирование методических рекомендаций по интеграции цифрового сервиса в учебный процесс, направленных на развитие критического отношения к информации в современном русскоязычном информационном пространстве.

Материалы и методы исследования. Антифейк-сервис основан на ансамблевом методе обучения нейросетевых моделей: рекуррентные нейронные сети (включая архитектуру нейронной сети долгой краткосрочной памяти), сверточные нейронные сети. Обучение проводилось на публичном наборе данных русскоязычных новостей “Fake and real Russian news”. Для предобработки текста на русском языке реализован специализированный конвейер, включающий лемматизацию и удаление стоп-слов. Оценка качества бинарной классификации новостных текстов на русском языке выполнялась на основе комплекса метрик: Accuracy, Precision, Recall и F1-score. Архитектура цифрового сервиса реализована с применением микросервисного подхода на базе технологического стека FastAPI/ React, что обеспечивает масштабируемость и гибкость системы.

Результаты исследования. Разработан рабочий прототип интеллектуального антифейк-сервиса для анализа русскоязычных новостных текстов. Ключевой особенностью предложенного цифрового решения является предоставлению пользователю возможности интерактивного выбора и комбинирования моделей глубокого обучения в реальном режиме времени через интуитивно

понятный web-интерфейс, что позволяет не только классифицировать новости на достоверные и фейковые, но и наглядно демонстрировать работу алгоритмов глубокого обучения в учебном процессе, формируя практические навыки верификации информации. Экспериментальные данные подтвердили эффективность предложенного подхода к разработке антифейк-сервиса: наивысшая точность (Accuracy 93,3%) была достигнута с использованием сверточной нейронной сети, что обусловлено ее способностью выявлять локальные семантические паттерны в текстах. Применение вероятностного ансамблирования повысило надежность и устойчивость бинарной классификации русскоязычных новостных текстов.

Заключение. Проведенное исследование решает актуальную проблему на стыке информационной безопасности и педагогики — противодействие дезинформации через развитие цифровой грамотности. Новизна и ключевое преимущество разработанного антифейк-сервиса заключаются в его функциональной двойственности: это инструмент автоматической бинарной классификации русскоязычных новостных текстов и одновременно сервис для формирования практических навыков верификации информации у пользователей через интерактивное взаимодействие с нейросетевыми моделями. Антифейк-сервис обладает значительным дидактическим потенциалом и может быть интегрирован в образовательные программы курсов для развития цифровой грамотности обучающихся, включая его использование в курсах по информатике, информационной безопасности, машинному обучению и другим дисциплинам. Таким образом, разработанный антифейк-сервис выполняет роль как инструмента автоматической верификации русскоязычного контента, так и цифрового средства образовательного назначения, которое способствует формированию критического отношения к информации в цифровой медиасреде.

Ключевые слова: антифейк-сервис, фейк, обработка естественного языка, глубокое обучение, нейронные сети, бинарная классификация, цифровая трансформация образования, цифровая грамотность.

Natalya A. Moiseeva¹, Arseniy D. Sotnikov²

¹ Omsk State Technical University, Omsk, Russia

² Seven bits, Omsk, Russia

Developing of an Anti-Fake Service of Educational Purposes for the Analysis of Russian-Language News Texts

Purpose of the research. The relevance of this study is driven by the steady growth of disinformation in the digital environment, which makes the development of automated content verification systems a critical tool for ensuring national information security. In addition, such systems serve as important digital technologies for educational purposes, helping to improve the digital literacy of modern users. The aim of this paper is to develop and substantiate the architecture of an anti-fake service as a specialized digital tool for education. The study addresses the following objectives: a comparative assessment of the effectiveness

of neural network models for analyzing news in Russian and the development of methodological recommendations for integrating digital services into the educational process to foster a critical attitude toward information in the contemporary Russian-language information space.

Materials and methods. The anti-fake service is based on an ensemble method for training neural network models, combining recurrent neural networks (including long short-term memory architecture) and convolutional neural networks. The training was conducted on the public data set of Russian-language news “Fake and real Russian

news". A specialized preprocessing pipeline, including lemmatization and stop word removal, was implemented for Russian-language text. The quality of binary classification for the Russian-language news was assessed using a set of metrics: Accuracy, Precision, Recall, and F1-score. The digital service architecture is implemented using a microservices approach based on the FastAPI/ React technology stack, ensuring system scalability and flexibility.

The results. It is developed a working prototype of an intelligent anti-fake service for analyzing Russian-language news texts. A key feature of the proposed digital solution is that it allows users to interactively select and combine deep learning models in real time via an intuitive web interface. This functionality not only enables news classification as reliable or fake, but also clearly demonstrates the inner workings of deep learning algorithms in an educational context, helping users to develop practical information verification skills. Experimental data confirm the effectiveness of the proposed approach to developing an anti-fake service: the highest accuracy (Accuracy 93.3%) was achieved using a convolutional neural network, owing to its ability to identify local semantic patterns in texts. The use of a probabilistic ensemble further improved the reliability and robustness of binary classification of Russian-language news.

Conclusion. The conducted research solves a pressing problem at the intersection of information security and pedagogy — countering disinformation through the development of digital literacy. The novelty and key advantage of the developed anti-fake service lie in its functional duality: it serves both as a tool for the automatic binary classification of Russian-language news texts and at the same time, as a service for developing users' practical information verification skills through interactive engagement with neural network models. The anti-fake service has significant didactic potential and can be integrated into educational programs for courses to develop students' digital literacy, including its use in courses on computer science, information security, machine learning, and other disciplines. Thus, the developed anti-fake service functions both as a tool for automatic verification of Russian-language content and as a digital educational tool that promotes the formation of a critical attitude toward information in the digital media environment.

Keywords: anti-fake service, fake, natural language processing; deep learning; neural networks; binary classification; digital transformation of education; digital literacy.

Введение

Современный этап развития цифровых технологий характеризуется не только ростом доступности информации, но и обострением проблемы ее достоверности. Широкое распространение социальных сетей и медийных платформ привело к увеличению объемов дезинформации во всем мире, в том числе фейковых новостей. «Фейковые новости» представляют собой «сфабрированную» информацию, которая внешне имитирует контент новостных СМИ, но не соответствует их организационным процессам или намерениям. В отличие от традиционных новостных агентств, фейковые новостные источники не придерживаются редакционных стандартов и процессов, направленных на обеспечение точности и достоверности информации. Фейковые новости пересекаются с другими видами информационных нарушений, такими как дезинформация (ложная информация, распространяемая с намерением обмануть людей) [1]. Для русскоязычного сегмента Internet эта проблема приобретает особую актуальность: согласно исследованию медиахолдинга Rambler&Co [2], проведенному среди российских Internet-пользователей в 2023 г., более 90% респондентов

сталкивались с фейковыми новостями, причем основным каналом их распространения (64%) являются социальные сети и мессенджеры. В условиях, когда уровень доверия к новостям в мире остается низким (по данным отчета “Edelman Trust Barometr 2025” [3] лишь 42% людей доверяют большинству новостей), традиционные методы верификации (фактчекинг, экспертный анализ) не могут обеспечить оперативную обработку постоянно растущих массивов данных. Все вышесказанное обуславливает объективную потребность в «умных» цифровых решениях, способных осуществлять первичную фильтрацию и оценку достоверности контента.

В ответ на этот вызов в последние годы активно развиваются фактчекинговые системы или системы детекции фейковых новостей (Fake News Detection, FND) [4–7], основанные на алгоритмах машинного обучения (Machine Learning, ML) и глубокого обучения (Deep Learning, DL), которые доказали свою эффективность в прикладном анализе текстов [8, 9]. Однако большинство существующих исследований и разработок ориентированы на англоязычный контент, в то время как лингвистические и культурные особенности русского языка создают

специфические сложности для автоматического анализа. В России, к примеру, существует лишь один фактчекинговый сервис — «Лапша» [4]. Параллельно с технологическим развитием возникает острая образовательная потребность: в эпоху цифровой трансформации развитие цифровой грамотности, включающей навыки критической оценки информации [10], становится одной из ключевых компетенций современного человека. Несмотря на это, в образовательной практике наблюдается дефицит инструментов, которые позволяли бы не только идентифицировать фейковые новости, но и наглядно демонстрировать процесс их выявления, что особенно важно для обучения работе с информацией в русскоязычной цифровой среде.

Для комплексного решения указанных задач в статье предлагается интегрированный подход, сочетающий технологический и образовательный аспекты. Разрабатываемый антифейк-сервис образовательного назначения представляет собой web-приложение, построенное на ансамбле моделей DL, решающий двудединую задачу: автоматическая бинарная классификация русскоязычных новостных текстов на достоверные и недостоверные,

а также формирование у пользователей практических навыков верификации информации через интерактивное взаимодействие с нейросетевыми моделями. Цель статьи: разработать и обосновать архитектуру антифейк-сервиса как цифрового образовательного инструмента, провести сравнительную оценку эффективности нейросетевых моделей для анализа русскоязычных новостей и сформулировать методические рекомендации по его интеграции в учебный процесс для развития цифровой грамотности и критического отношения к информации в современном информационном пространстве.

1. Анализ современных алгоритмов искусственного интеллекта и доступных наборов данных для разработки антифейк-сервиса

Прежде чем перейти к описанию предлагаемого решения, необходимо проанализировать современное состояние исследований в области FND. Такой анализ позволит выявить ключевые тенденции, используемые методы и, что особенно важно для данного исследования, определить существующие пробелы, связанные с работой именно с русскоязычным контентом.

В последние годы активно развиваются подходы к FND, основанные на алгоритмах ML. Например, в работе [7] была предложена система, использующая метод обратного отслеживания для оценки вероятности наличия фейковых новостей, которая продемонстрировала производительность свыше 85%. Другие исследователи сосредоточились на выявлении поддельных отзывов в Internet-магазинах, установив высокую эффективность классификатора AdaBoost [11]. В рамках анализа закономерностей распространения информации приме-

нялись статистические методы извлечения признаков из текстовых данных: TF-IDF (Term Frequency-Inverse Document Frequency) для оценки важности терминов и латентное распределение Дирихле (Latent Dirichlet Allocation, LDA) для тематического моделирования новостного потока в социальных медиа [12]. Кроме того, вычислительно-стилистическая оценка на основе статистических методов обработки естественного языка (Natural Language Processing, NLP) показала свою эффективность для FND в текстах из социальных сетей [13]. Эти работы подтверждают потенциал ML-подходов в решении задач верификации контента.

Параллельно с методами ML стремительно развиваются подходы на основе DL. В 2019 г. F. Özbay и B. Alatas предложили двухэтапную стратегию, сочетающую предобработку текста и интеллектуальный анализ данных с помощью алгоритмов ML [14]. В том же году S.S. Jadhav и S.D. Thepade достигли точности в 99%, применив модель Deep Structure Semantic Model (DSSM) и рекуррентные нейронные сети (Recurrent Neural Networks, RNN) [15]. Дальнейшее развитие направления связано с созданием ансамблевых моделей, таких как многоуровневая система голосования с использованием двенадцати классификаторов ML [16], а также применением специализированных архитектур: оптимизированной сверточной нейронной сети (Convolutional Neural Network, CNN) для выявления фейков, а именно OPCNN-FAKE [17], моделей на основе долгой краткосрочной памяти (Long Short-Term Memory, LSTM) [18, 19], фреймворк ConvNet для многомодального обнаружения фейковых новостей [20] и гибридных подходов (CNN-LSTM с логистической регрессией) [21]. Высокую эффективность также продемон-

стрировали трансформерные архитектуры, такие как BERT, RoBERTa и XLNet [22].

Однако подавляющее большинство перечисленных исследований ориентировано на анализ англоязычного контента. В то время как проблема фейковых новостей носит глобальный характер, лингвистические, стилистические и культурные особенности русского языка создают специфические сложности для автоматического анализа, что требует адаптации существующих и разработки новых методов. Аналитический обзор работ, посвященных непосредственно русскоязычным новостям, выявил их сравнительно малое количество, что указывает на значительный пробел в данной области. Например, в исследовании [23] разработана нейронная сеть типа персептрон, обучаемая при работе web-краулера для выявления потенциально ложных статей. В работе [24] методы ML применялись для детекции фейков в социальных сетях. Более современные подходы, представленные в [25], предлагают гибридные архитектуры на основе BERT (BERT-CNN, BERT-GTU, BERT-LSTM). Тем не менее, эти исследования носят скорее фрагментарный характер и не формируют целостной картины относительно доступных и эффективных методов для русского языка.

Важно отметить, что алгоритмы ML и DL требуют для своей работы качественных и репрезентативных данных. Существенной проблемой в области FND является отсутствие крупномасштабных и точно размеченных эталонных наборов данных, особенно для русскоязычного сегмента. Исследователи часто вынуждены использовать узкоспециализированные или самостоятельно собранные коллекции текстов, что затрудняет сравнение и воспроизведение результатов.

Учитывая цель данной работы — создание антифейк-сервиса, адаптированного для русскоязычной среды, — в качестве базового был выбран общедоступный датасет “Fake and real Russian news”, который содержит поддельные и реальные новости на русском языке, что позволяет проводить адекватное обучение и валидацию моделей для целевой языковой среды.

Таким образом, проведенный анализ подтверждает, что, несмотря на активное развитие методов FND, для русскоязычного сегмента эта область остается недостаточно изученной. Существуют исследования немногочисленные, а проблема отсутствия качественных размеченных данных усугубляет ситуацию. Это подчеркивает практическую значимость настоящей работы, направленной на создание комплексного решения — антифейк-сервиса, который сочетает адаптированные модели DL с учебно-методическим функционалом для решения актуальной проблемы дезинформации в русскоязычном цифровом пространстве.

2. Разработка модели глубокого обучения для детекции фейков в новостях на русском языке

На основе анализе существующих подходов к FND, представленного в предыдущем разделе, была разработана модель DL, адаптированная для работы с русскоязычным контентом. Важно подчеркнуть, что ядром функционала интеллектуального антифейк-сервиса является алгоритм бинарной классификации текстовых данных, реализованный на основе ансамбля моделей DL: RNN, LSTM и CNN. Такой выбор позволяет комплексно подойти к анализу текста, сочетая способности моделей улавливать как последовательные зависимости, так и локальные семантические паттерны. Об-

щая схема работы разработанного алгоритма состоит из двух ключевых этапов (рис. 1): предобработка текста и классификации с помощью нейросетевых моделей.

Формальная постановка задачи. Задача FND формализуется как задача бинарной классификации с учителем. Для выборки из N текстов, где каждый текст представлен парой X_i и y_i , необходимо определить, относится ли новость к классу 0 (фейк, False) или 1 (достоверная, True):

$$D = \{(X_1, y_1), (X_2, y_2), \dots, (X_N, y_N)\}, \quad (1)$$

где X_i — вектор признаков из текста новости, $y_i \in \{0, 1\}$ — метка класса.

Цель обучения с учителем: найти функцию прогноза $f(X; W)$, которая аппроксимирует истинное распределение меток:

$$f(X_i; W) \approx P(y_i = 1 | X_i), \quad (2)$$

где $f(X_i; W)$ возвращает вероятность того, что новость является достоверной.

Решение принимается по пороговому правилу с $\tau = 0,5$:

$$\hat{y}_i = \begin{cases} 1, & \text{если } f(X_i, W) \geq \tau \\ 0, & \text{иначе} \end{cases}, \quad (3)$$

где $\tau \in (0, 1)$ — порог классификации, обычно равный 0,5.

Для обучения и оценки моделей в исследовании был использован общедоступный датасет “Fake and real Russian news”, содержащий 5497 записей (примерно поровну достоверных и фейковых новостей). Выбор этого датасета был обусловлен его релевантностью целевой языковой среде и сбалансированностью. Данные были разделены на обучающую (56%), валидационную (24%) и тестовую (20%) выборки для обеспечения надежности. Для обучения использовалась функции потерь бинарная кросс-энтропия — стандартный выбор для задач бинарной классификации [21]. Эта функция оценивает рас-

хождение между предсказанным значением вероятности и истинной меткой класса:

$$L(y, \hat{y}) = -(y * \log(\hat{y}) + (1 - y) * \log(1 - \hat{y})), \quad (4)$$

где $y \in \{0, 1\}$ — истинная метка класса, а $\hat{y} \in (0, 1)$ — предсказанная моделью вероятность принадлежности к положительному классу. Данная функция является дифференцируемой и хорошо подходит для оптимизации при помощи градиентных методов, обеспечивая плавное обновление параметров модели.

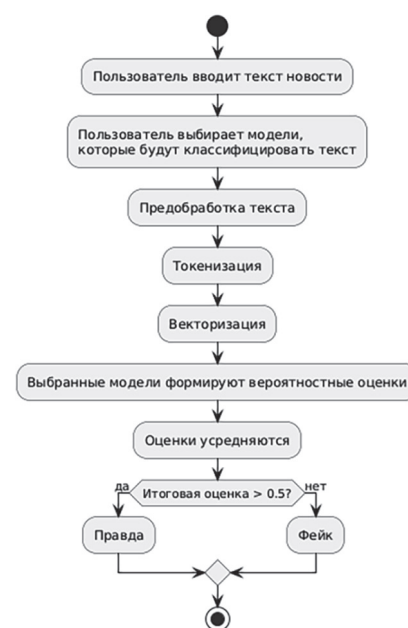


Рис. 1. Диаграмма, иллюстрирующая процесс классификации русскоязычного новостного текста

Fig. 1. Diagram illustrates the process of classifying Russian-language news text

На выходном слое каждой нейронной сети применялась сигмоидальная функция активации:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (5)$$

где z — входное значение нейрона.

Сигмоида преобразует любое вещественное число в диапазон $(0, 1)$, что позволяет интерпретировать результат как вероятность принадлежно-

сти к классу. Её использование оправдано в задачах бинарной классификации, поскольку она естественным образом сочетается с функцией бинарной кросс-энтропии, позволяя обучать модель с вероятностным выходом.

Критически важным этапом, особенно для флективного русского языка, стала специализированная предобработка текстов. Для каждого примера текст формировался путем конкатенации заголовка и основного содержания статьи, что позволило объединить краткую и контекстную информацию. Далее выполнялся специализированный NLP-конвейер, адаптированный под морфологические особенности русского языка: 1) Приведение к нижнему регистру для устранения вариативности написания. 2) Токенизация с учетом морфологических особенностей. 3) Удаление стоп-слов (использовался список стоп-слов библиотеки NLTK для русского языка), знаков препинания и неинформативных токенов. 4) Лемматизация с использованием модели `ru_core_news_sm` из NLP-библиотеки `spaCy` для приведения слов к их нормальной (словарной) форме. Этот этап особенно важен для флективного русского языка, т. к. позволяет устранить морфологические вариации (падежи, склонения, спряжения) и значительно снизить размерность признакового пространства.

В результате предобработки каждый текст представлялся в виде списка лемм. Анализ показал, что средняя длина текста после обработки составляет 142 токенов, а 95% словоупотреблений покрываются словарем в 10 000 наиболее частотных слов, что и было установлено в качестве максимального размера словаря. Затем проводилась векторизация, которая заключалась в преобразовании токенов в целочисленные индексы из

Сравнение алгоритмов глубокого обучения
Comparison of deep learning algorithms

Модель	Точность	Precision (Fake)	Precision (True)	Recall (Fake)	Recall (True)
CNN	0,9464	0,92	0,95	0,95	0,92
LSTM	0,8946	0,86	0,93	0,93	0,86
RNN	0,8783	0,87	0,89	0,88	0,88

частотного словаря, все последовательности приводились к единой длине в 150 токенов с помощью дополнения нулями или усечения.

Были разработаны, обучены и протестированы три нейросетевые архитектуры (рис. 3). Все модели обучались с использованием оптимизатора Adam и функции потерь бинарной кросс-энтропии – стандартного выбора для задач бинарной классификации, обеспечивающего плавное обновление параметров. На выходном слое применялась сигмоидальная функция активации для получения вероятностной оценки.

Работа системы организована как интерактивный процесс (см. рис. 1). Пользователь через web-интерфейс

вводит текст новости и выбирает одну или несколько моделей для анализа (RNN, LSTM, CNN). Каждая модель возвращает вероятность принадлежности текста к классу «достоверных» новостей. Итоговая оценка вычисляется как среднее арифметическое вероятностей от выбранных моделей (метод усреднения, или «мягкое» голосование). Такой ансамблевый подход позволяет повысить надежность и устойчивость системы, снижая влияние ошибок, присущих каждой отдельной модели. Окончательное решение (фейк/ достоверно) принимается на основе порога $\tau = 0,5$, после чего результат сохраняется в базу данных (рис. 2) и отображается пользователю. Для орга-

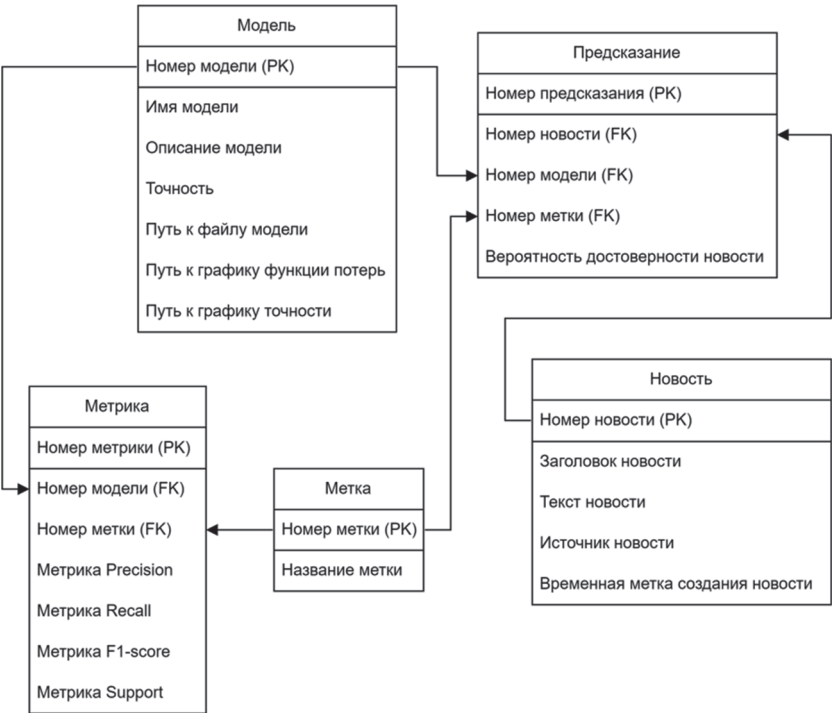


Рис. 2. Логическая модель базы данных “AntiFake”
Fig. 2. Logical model of the “AntiFake” database

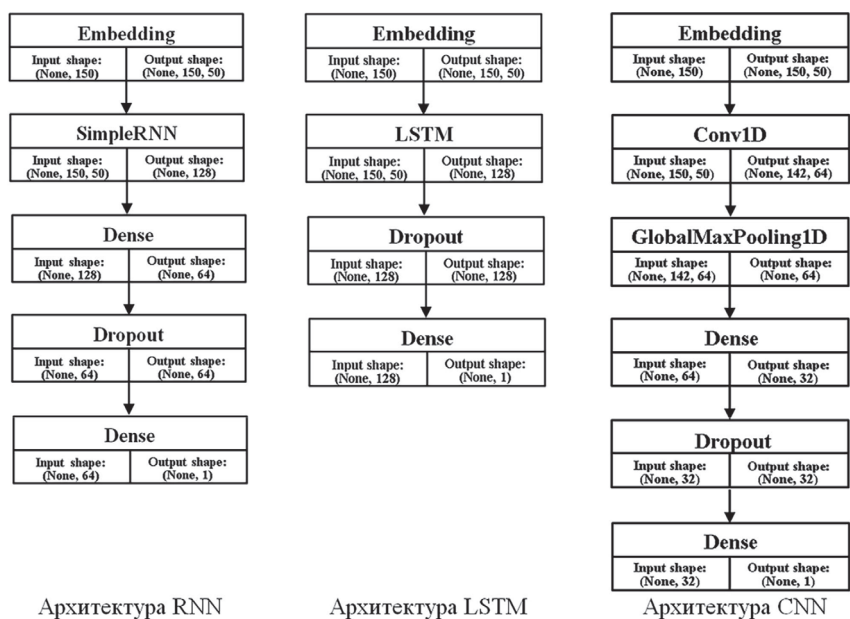


Рис. 3. Архитектуры нейронных сетей для классификации новостных фейков на русском языке

Fig. 3. Neural network architectures for classifying fake news in Russian

низации и хранения данных в антифейк-сервисе используется реляционная база данных (БД), т.к. она хорошо подходит для работы со структурированными данными, которые могут быть организованы в таблицы.

Для оценки качества классификации использовался комплекс метрик: Accuracy, Precision, Recall и F1-score, что особенно важно для задач, где ложные срабатывания имеют критическое значение [26]. Сравнительный анализ эффективности моделей (см. рис. 3) демонстрирует существенные различия в их работе. Результаты на тестовой выборке (см. табл.) подтвердили сбалансированности классификации по обоим классам и выявили существенные различия в эффективности моделей DL.

RNN показывает базовую эффективность (**Accuracy 87,8%**), но проявила склонность к проблеме затухающих градиентов при обработке длинных последовательностей. В отличие от нее, LSTM, благодаря механизму ворот (гейтинг, Gate), эффективнее работала с долгосрочными зависимостями, что повысило точность до **89,5%**. Наилучший результат показывает CNN (Accuracy 93,3%), что объясняется ее способностью выявлять локальные семантические и стилистические паттерны благодаря использованию ядра свертки размером 9. Ключевое преимущество CNN – эффективная идентификация специфических лингвистических маркеров фейковых новостей, характерных для русскоязычного контента. Например, такие расплывчатые формулировки типа «Эксперты утверждают...» без указания фамилий экспертов или конкретных источников, часто свидетельствует о манипулятивном характере контента.

Таким образом, итоговый алгоритм FND реализует следующий NLP-конвейер: «предобработка текста → ансамблевая классификация» (см. рис. 1). Наилучшие результаты показала архитектура CNN, что делает её наиболее предпочтительной для бинарной классификации русскоязычных новостей. Практическая ценность разработанного решения заключается не только в высокой точности, но и в возможности гибкого выбора и комбинирования моде-

лей пользователем, а также в обеспечении специализированного конвейера предобработки для русского языка, что формирует эффективный и адаптируемый инструмент для противодействия дезинформации в русскоязычном цифровом пространстве.

3. Архитектура интеллектуального антифейк-сервиса

Разработанный алгоритм DL для FND на русском языке был реализован в форме полнофункционального web-приложения. Его архитектура была спроектирована с учетом двух ключевых требований: технологической эффективности (обеспечение высокой производительности и точности классификации) и образовательной направленности (обеспечение интерактивности, наглядности и простоты интеграции в учебный процесс). Для достижения этих целей была выбрана микросервисная клиент-серверная архитектура. Такой подход обеспечивает:

- Масштабируемость: возможность независимого масштабирования компонентов (например, сервиса предобработки текста или модуля классификации) при росте нагрузок.
- Гибкость и удобство поддержки: изолированные сервисы проще разрабатывать, тестировать и обновлять, что особенно важно для исследовательского и образовательного инструмента, требующего постоянного улучшения.
- Надежность: отказ одного из микросервисов не приводит к полной неработоспособности системы.
- Технологический стек был подобран в соответствии с современными стандартами web-разработки и задачами ML:
- бэкенд-разработка: высокоуровневый язык программирования Python, веб-фреймворк FastAPI;

- фронтенд-разработка: высокоуровневый язык программирования TypeScript и web-фреймворк React;
- обучение и развертывание моделей DL (CNN, LSTM, RNN): фреймворк TensorFlow с высокоуровневым API-интерфейсом Keras;
- хранение данных: реляционная система управления базами данных MySQL.

Общая схема архитектуры приложения представлена на рисунке 4, которая визуализирует все компоненты системы (клиент, сервер, БД, DL-модели) и потоки данных между ними, отражая логическую структуру интеллектуального цифрового решения для FND на русском языке. Основу системы составляет алгоритм бинарной классификации текстов, реализованный на основе трех нейросетевых архитектур: CNN, LSTM и RNN. Ключевой особенностью является возможность выбора ансамбля моделей пользователем, что позволяет сравнивать их эффективность и выбирать подходящую для конкретной задачи. Рассмотрим ключевые компоненты системы (см. рис. 4).

• Клиентское приложение (Frontend) реализовано на React и предоставляет пользовательский интерфейс. Его функционал включает: 1) ввод и загрузку текста новости; 2) интерактивный выбор одной или нескольких моделей DL (CNN, LSTM, RNN) для анализа; 3) наглядное отображение результатов классификации: вероятностные оценки, результат анализа текста новости на русском языке, а также возможность сравнения результатов разных моделей DL; 4) отправка запросов и получение данных от серверной части через REST API.

• Серверная часть (Backend) реализована на FastAPI и состоит из нескольких логических модулей: 1) REST API — единая точка входа для всех

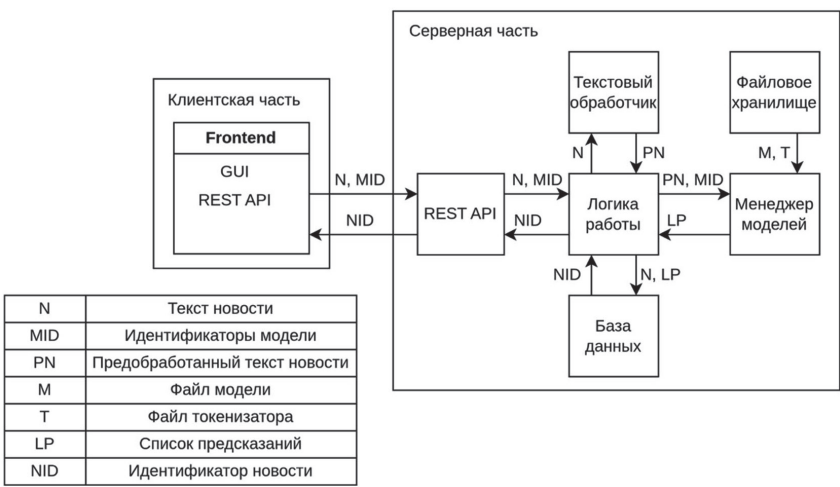


Рис. 4. Схема архитектуры приложения “АнтиФейк”
Fig. 4. “AntiFake” application architecture diagram

клиентских запросов; 2) текстовый обработчик выполняет критически важный этап предобработки входящего текста. Этот модуль инкапсулирует всю логику работу с морфологией русского языка.

• Менеджер моделей отвечает за загрузку обученных моделей DL из файлового хранилища, управление их жизненным циклом в памяти и выполнения предсказаний на подготовленных данных.

• Логика работы координирует весь рабочий процесс: принимает запрос от API, передает текст на предобработку, отправляет векторизованные данные выбранным моделям DL через компонент менеджера моделей, агрегирует их результаты (усреднение вероятностей), применяет пороговое правило и формирует итоговый ответ.

Таким образом, реализованная архитектура антифейк-сервиса представляет собой масштабируемое и эффективное цифровое решение, где каждый компонент выполняет строго определенную функцию. Ядро системы на основе DL демонстрирует высокую точность в решении задачи бинарной классификации новостных текстов, что подтверждается результатами экспериментальных исследований.

4. Образовательный потенциал антифейк-сервиса и его применение в учебном процессе

Разработанный антифейк-сервис представляет собой не только технологический инструмент для детекции фейков, но и инновационный цифровой ресурс образовательного назначения, созданный для решения дефицита практико-ориентированных средств обучения в области цифровой грамотности. Его архитектура и функционал изначально проектировались с учетом дидактических задач, что позволяет комплексно интегрировать его в образовательный процесс. Рассмотрим ключевые образовательные функции системы:

1) Интерактивное изучение принципов работы алгоритмов DL при анализе новостных текстов на русском языке.

Антифейк-сервис позволяет пользователям (обучающимся) в интерактивном режиме загружать новостные тексты, выбирать для их анализа одну или несколько нейросетевых моделей (CNN, LSTM, RNN) и сравнить полученные результаты классификации. Это дает наглядное представление о работе алгоритмов DL и их способности выявлять лингвистические маркеры фейковых

новостей на русском языке.

2) Формирование практических навыков верификации информации (фактчекинга).

Используя антифейк-сервис для проверки текстов в режиме, приближенному к реальному времени, пользователи развивают навыки критического анализа и оценки информации: учатся выявлять сомнительные формулировки в источниках информации, проверять контекст и сопоставлять данные, что составляет основу цифровой грамотности

3) Интеграция в учебные программы и цифровые образовательные платформы.

Благодаря модульной микросервисной архитектуре и наличию REST API, антифейк-сервис обладает высокой степенью интеграционной гибкости. Он может быть интегрирован в:

- системы управления обучением (Learning Management System, LMS, например, Moodle) в виде внешнего интеллектуального цифрового инструмента для проведения лабораторных работ по цифровой грамотности с помощью стандарта LTI (Learning Tools Interoperability, совместимость учебных инструментов), разработанного для облегчения взаимодействия LMS с различными цифровыми образовательными ресурсами и цифровыми сервисами [27];

- цифровые образовательные платформы и онлайн-курсы по таким дисциплинам как информационная безопасность, компьютерная лингвистика, основы ML и DL, реализуемые в формате лабораторных работ, практикумов и междисциплинарных проектов;

- междисциплинарные проекты, сочетающие элементы обществознания, социологии, информатики, русского языка и культуры речи.

4) Создание БД учебных кейсов для анализа.

Все поступающие на анализ тексты и результаты их клас-

сификации автоматически сохраняются в БД, что позволяет преподавателю:

- формировать тематические коллекции реальных примеров (кейсов) — как фейковых, так и достоверных новостей — для детального разбора на занятиях;

- отслеживать эволюцию типичных маркеров дезинформации и адаптировать учебный материал под конкретные актуальные задачи, выявлять информационные угрозы;

- использовать накопленные данные для новых исследовательских задач (например, для разметки датасетов следующего поколения).

5) Поддержка проектной и исследовательской деятельности обучающихся.

Антифейк-сервис предоставляет технологическую основу для выполнения учебно-исследовательских проектов, позволяя обучающимся фокусироваться на аналитической, а не на технической реализации. Примеры возможных тем проектов:

- Сравнительный анализ эффективности различных архитектур нейронных сетей на новостях разной тематики (экономика, образование, здоровье и т.д.).

- Исследование региональных особенностей дезинформации на основе самостоятельно собранных новостных датасетов;

- Лингвистический анализ специфических конструкций, характерных для недостоверного русскоязычного контента.

Таким образом, антифейк-сервис выступает не только в качестве цифрового интеллектуального инструмента для верификации контента, способствуя повышению уровня цифровой грамотности и устойчивости к дезинформации в русскоязычной цифровой среде, но и интегрируется в учебный процесс. Его внедрение позволяет минимизировать разрыв между теоре-

тическими знаниями и практическими навыками работы с цифровыми технологиями в условиях информационной неопределенности, вызванной цифровизацией общества.

Заключение

Проведенное исследование посвящено решению актуальной проблемы, находящейся на стыке информационной безопасности и педагогической практики, — противодействие дезинформации через развитие цифровой грамотности. В рамках работы разработан прототип интеллектуального антифейк-сервиса образовательного назначения для анализа русскоязычных новостных текстов. Web-приложение реализовано на основе микросервисной архитектуры (FastAPI/React) и использует ансамблевый подход, объединяющий три нейросетевые архитектуры (CNN, LSTM, RNN), что позволило не только создать рабочий цифровой инструмент классификации, но и провести сравнительную оценку эффективности моделей DL. Применение метода усреднения вероятностей («мягкое» голосование) в ансамбле позволило повысить надежность и устойчивость системы по сравнению с использованием одиночных классификаторов.

Экспериментальные результаты, полученные на публичном датасете русскоязычных новостей, подтвердили эффективность предложенного подхода: наивысшая точность (Accuracy 93,3%) показала CNN-архитектура, превзойдя результаты LSTM (89,5%) и RNN (87,8%). Это объясняется способностью CNN эффективно выявлять локальные семантические и стилистические паттерны, характерные для фейковых новостей на русском языке. Ключевым преимуществом разработанного антифейк-сервиса является его двойная направленность: он

служит как цифровой инструмент автоматической верификации, так и интерактивного сервиса образовательного назначения. Ансамблевый подход, предоставляющий пользователям возможность выбора и комбинирования моделей DL через интуитивно понятный web-интерфейс, делает процесс анализа прозрачным и адаптивным, что особенно ценно для использования в учебном процессе.

Антифейк-сервис обладает значительным дидактическим потенциалом и может быть интегрирован

в образовательные программы курсов по информатике, цифровой грамотности, информационной безопасности, ML, а также использован для самообразования и повышении квалификации педагогов в области цифровых технологий. Основные идеи и результаты работы были апробированы на XV Международной научно-практической конференции [28]. Перспективы дальнейших исследований связаны с развитием функционала антифейк-сервиса: внедрение более современных архитектур на основе трансформеров

(например, BERT), поддержка мультязычных данных, разработка специализированных учебных модулей с системами заданий и инструментами аналитики для отслеживания прогресса обучаемых. Таким образом, данная работа вносит вклад как в область прикладной информатики для анализа текста, так и цифровую дидактику, предлагая практическое, апробированное решение для повышения уровня цифровой грамотности пользователей и противодействия дезинформации в русскоязычном цифровом пространстве.

Литература

1. Lazer D., Baum M., Benkler Y., Berinsky A., Greenhill K., Menczer F., Metzger M., Nyhan B., Pennycook G., Rothschild D., Schudson M., Sloman S., Sunstein C., Thorson E., Watts D., Zittrain J. The science of fake news // *Science*. 2018. Т. 359. № 6380. С. 1094–1096. DOI: 10.1126/science.aao2998.
2. Исследование: более 90% опрошенных россиян сталкивались с фейками [Электрон. ресурс] // *Rambler&Co*. 2023. Режим доступа: <https://rambler-co.ru/news/638>.
3. 2025 Edelman Trust Barometr. Global Report [Электрон. ресурс]. Режим доступа: https://www.edelman.com/sites/g/files/aatuss191/files/2025-01/2025%20Edelman%20Trust%20Barometer%20Global%20Report_01.23.25.pdf.
4. В России появится первый антифейк-сервис [Электрон. ресурс]. Режим доступа: <https://dialog.info/v-rossii-poyavitsya-pervyj-antifejk-servis/>.
5. Mahalpure A., Marwade A., Kadam P., Raut Sh. Advancing media integrity through AI-powered fake news detection // *International journal of research and analytical reviews*. 2025. Т. 12. № 2. DOI: 10.56975/ijrar.v12i2.310939.
6. Gifu D. An intelligent system for detecting fake news // *Procedia Computer Science*. 2023. Т. 221. С. 1058–1065. DOI: 10.1016/j.procs.2023.08.088.
7. Ko H., Hong J.Y., Kim S., Mesicek L., Na I.S. Human-machine interaction: a case study on fake news detection using a backtracking based on a cognitive system // *Cognitive Systems Research*. 2019. Т. 55. № 11. С. 77–81. DOI: 10.1016/j.cogsys.2018.12.018.
8. Бенгфорт Б., Билбро Р., Охеда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. СПб.: Питер, 2019. 368 с.
9. Sarkar D. Text analytics with python. A practitioner's guide to natural language processing. India: 2019, 674 с.
10. Баймуратова Л.Р., Долгова О.А., Имаева Г.Р., Гриценко В.И., Смирнов К.В. Цифровая грамотность для экономики будущего [Электрон. ресурс]. Режим доступа: <https://nafi.ru/projects/sotsialnoe-razvitie/tsifrovaya-gramotnost-dlya-ekonomiki-budushchego/>.
11. Barbado R., Araque O., Iglesias C.A. A framework for fake review detection in online consumer electronics retailers // *Information Processing and Management*. 2019. Т. 56. № 4. С. 1234–1244. DOI: 10.1016/j.ipm.2019.03.002.
12. Xu K., Wang F., Wang H., Yang Bo. Detecting fake news over online social media via domain reputations and content understanding // *Tsinghua Science and Technology*. 2020. Т. 25. № 1. С. 20–27. DOI: 10.26599/tst.2018.9010139.
13. Oliveira N.R., Medeiros D.S.V., Mattos D.M.F. A sensitive stylistic approach to identify fake news on social networking // *IEEE Signal Processing Letters*. 2020. Т. 27. С. 1250–1254. DOI: 10.1109/lsp.2020.3008087.
14. Özbay F.A., Alatas B. Fake news detection within online social media using supervised artificial intelligence algorithms // *Physica A: Statistical Mechanics and its Applications*. 2020. Т. 540. С. 123174. DOI: 10.1016/j.physa.2019.123174.
15. Jadhav S.S., Thepade S.D. Fake news identification and classification using DSSM and improved recurrent neural network classifier // *Applied Artificial Intelligence*. 2019. Т. 33. № 12. С. 1058–1068. DOI: 10.1080/08839514.2019.1661579.
16. Kaur S., Kumar P., Kumaraguru P. Automating fake news detection system using multi-level voting model // *Soft Computing*. 2020. Т. 24. С. 9049–9069. DOI: 10.1007/s00500-019-04436-y.
17. Saleh H., Alharbi A., Alsamhi S.H. OPCNN-FAKE: Optimized convolutional neural

network for fake news detection // IEEE Access. 2021. Т. 9. С. 129471–129489. DOI: 10.1109/access.2021.3112806.

18. Bhattacharjee R., Sufian M., Talha M., Ahmed M., Tasnim F., Sara S., Islam K. Title-based fake news detection using LSTM. 2024. С. 173–178. DOI: 10.1109/ICDT61202.2024.10488933.

19. Khare S., Singh P., Kumar P. LSTM based approach to detect fake news. 2024. С. 1501–1505. DOI: 10.1109/ICDT61202.2024.10488933.

20. Raj Ch., Meel P. ConvNet frameworks for multi-modal fake news detection // Applied Intelligence. 2021. Т. 51. № 11. С. 8132–8148. DOI: 10.1007/s10489-021-02345-y.

21. Pummy D., Amandeep K. Automatic fake news identification: a hybrid CNN-LSTM-logistic regression approach // Journal of Information Systems Engineering and Management. 2025. Т. 10. № 7. С. 446–460. DOI: 10.52783/jisem.v10i7s.876.

22. Saadi A., Belhadeh H., Guessas A., Hafirassou O. Enhancing fake news detection with transformer models and summarization // Engineering, Technology and Applied Science Research. 2025. Т. 1. № 3. С. 23253–23259. DOI: 10.48084/etasr.10678.

23. Некрасов Г.А., Романова И.И. Разработка поискового робота для обнаружения веб-контента с фейковыми новостями // Инновационные, информационные и коммуникационные технологии. 2017. № 1. С. 128–130.

24. Жук Д.А., Жук Д.В., Третьяков А.О. Методы определения поддельных новостей в социальных сетях с использованием машинного обучения // Информационные ресурсы России. 2018. № 3(163). С. 29–32.

25. Тумбинская М.В., Галиев Р.А. Идентификация фейк-новостей с помощью веб-ресурса на основе нейронных сетей // Программные продукты и системы. 2023. № 4. С. 590–599. DOI: 10.15827/0236-235X.144.590-599.

26. Roumeliotis K.I., Tselikas N.D., Nasiopoulos D.K. Fake news detection and classification: a comparative study of convolutional neural networks, large language models, and natural language processing models // Future Internet. 2025. Т. 17. № 1. С. 28. DOI: 10.3390/fi17010028.

27. Внедрение стандарта совместимости учебных инструментов (LTI) [Электрон. ресурс]. Режим доступа: <https://arcadia.spb.ru/studies/lti-for-elearning>.

28. Моисеева Н.А., Сотников А.Д. Сравнительный анализ архитектур нейронных сетей для детекции фейковых новостей // XV Международная молодежная научно-практическая конференция с элементами научной школы «Прикладная математика и фундаментальная информатика» (Омск, 19–24 мая 2025 года). Омск: Омский государственный технический университет, 2025. С. 114–115.

References

1. Lazer D., Baum M., Benkler Y., Berinsky A., Greenhill K., Menczer F., Metzger M., Nyhan B., Pennycook G., Rothschild D., Schudson M., Sloman S., Sunstein C., Thorson E., Watts D., Zittrain J. The science of fake news. Science. 2018; 359; 6380: 1094–1096. DOI: 10.1126/science.aao2998.

2. Issledovaniye: boleye 90% oproshennykh rossiyan stalkivalis' s feykami = Research: More than 90% of Russians surveyed have encountered fakes [Internet]. Rambler&Co. 2023. Available from: <https://rambler-co.ru/news/638>. (In Russ.)

3. 2025 Edelman Trust Barometr. Global Report [Internet]. Available from: https://www.edelman.com/sites/g/files/aatuss191/files/2025-01/2025%20Edelman%20Trust%20Barometer%20Global%20Report_01.23.25.pdf.

4. V Rossii poyavitsya pervyy antifeyk-servis = Russia to Launch First Anti-Fake Service [Internet]. Available from: <https://dialog.info/v-rossii-poyavitsya-pervyj-antifejk-servis/>. (In Russ.)

5. Mahalpure A., Marwade A., Kadam P., Raut Sh. Advancing media integrity through AI-powered fake news detection. International journal of research and analytical reviews. 2025; 12: 2. DOI: 10.56975/ijrar.v12i2.310939.

6. Gifu D. An intelligent system for detecting fake news. Procedia Computer Science. 2023; 221: 1058–1065. DOI: 10.1016/j.procs.2023.08.088.

7. Ko H., Hong J.Y., Kim S., Mesicek L., Na I.S. Human-machine interaction: a case study on fake news detection using a backtracking based on a cognitive system. Cognitive Systems Research. 2019; 55; 11: 77–81. DOI: 10.1016/j.cogsys.2018.12.018.

8. Bengfort B., Bilbro R., Okheda T. Prikladnoy analiz tekstovoykh dannykh na Python. Mashinnoye obucheniye i sozdaniye prilozheniy obrabotki yestestvennogo yazyka = Applied Text Analytics with Python. Machine Learning and Building Natural Language Processing Applications. Saint Petersburg: Piter; 2019. 368 p. (In Russ.)

9. Sarkar D. Text analytics with python. A practitioner's guide to natural language processing. India: 2019; 674 p.

10. Baymuratova L.R., Dolgova O.A., Imayeva G.R., Gritsenko V.I., Smirnov K.V. Tsifrovaya gramotnost' dlya ekonomiki budushchego = Digital Literacy for the Economy of the Future [Internet]. Available from: <https://nafi.ru/projects/sotsialnoe-razvitie/tsifrovaya-gramotnost-dlya-ekonomiki-budushchego/>. (In Russ.)

11. Barbado R., Araque O., Iglesias C.A. A framework for fake review detection in online consumer electronics retailers. Information Processing and Management. 2019; 56; 4: 1234–1244. DOI: 10.1016/j.ipm.2019.03.002.

12. Xu K., Wang F., Wang H., Yang Bo. Detecting fake news over online social media via domain reputations and content understanding.

Tsinghua Science and Technology. 2020; 25; 1: 20–27. DOI: 10.26599/tst.2018.9010139.

13. Oliveira N. R., Medeiros D.S.V., Mattos D.M.F. A sensitive stylistic approach to identify fake news on social networking. IEEE Signal Processing Letters. 2020; 27: 1250–1254. DOI: 10.1109/lsp.2020.3008087.

14. Özbay F.A., Alatas B. Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications. 2020; 540: 123174. DOI: 10.1016/j.physa.2019.123174.

15. Jadhav S.S., Thepade S.D. Fake news identification and classification using DSSM and improved recurrent neural network classifier. Applied Artificial Intelligence. 2019; 33; 12: 1058–1068. DOI: 10.1080/08839514.2019.1661579.

16. Kaur S., Kumar P., Kumaraguru P. Automating fake news detection system using multi-level voting model. Soft Computing. 2020; 24: 9049–9069. DOI: 10.1007/s00500-019-04436-y.

17. Saleh H., Alharbi A., Alsamhi S.H. OPCNN-FAKE: Optimized convolutional neural network for fake news detection. IEEE Access. 2021; 9: 129471–129489. DOI: 10.1109/access.2021.3112806.

18. Bhattacharjee R., Sufian M., Talha M., Ahmed M., Tasnim F., Sara S., Islam K. Title-based fake news detection using LSTM. 2024: 173–178. DOI: 10.1109/ICDT61202.2024.10488933.

19. Khare S., Singh P., Kumar P. LSTM based approach to detect fake news. 2024: 1501–1505. DOI: 10.1109/ICDT61202.2024.10488933.

20. Raj Ch., Meel P. ConvNet frameworks for multi-modal fake news detection. Applied Intelligence. 2021; 51; 11: 8132–8148. DOI: 10.1007/s10489-021-02345-y.

21. Pummy D., Amandeep K. Automatic fake news identification: a hybrid CNN-LSTM-logistic regression approach. Journal of Information Systems Engineering and Management. 2025; 10; 7: 446–460. DOI: 10.52783/jisem.v10i7s.876.

22. Saadi A., Belhadeh H., Guessas A., Hafirassou O. Enhancing fake news detection with

transformer models and summarization. Engineering, Technology and Applied Science Research. 2025; 1; 3: 23253–23259. DOI: 10.48084/etasr.10678.

23. Nekrasov G.A., Romanova I.I. Development of a search robot for detecting web content with fake news. Innovatsionnyye, informatsionnyye i kommunikatsionnyye tekhnologii = Innovative, information and communication technologies. 2017; 1: 128–130. (In Russ.)

24. Zhuk D.A., Zhuk D.V., Tret'yakov A.O. Methods for identifying fake news in social networks using machine learning. Informatsionnyye resursy Rossii = Information resources of Russia. 2018; 3(163): 29–32. (In Russ.)

25. Tumbinskaya M.V., Galiyev R.A. Identification of fake news using a web resource based on neural networks. Programmnyye produkty i sistemy = Software products and systems. 2023; 4: 590–599. DOI: 10.15827/0236-235X.144.590-599. (In Russ.)

26. Roumeliotis K.I., Tselikas N.D., Nasiopoulos D.K. Fake news detection and classification: a comparative study of convolutional neural networks, large language models, and natural language processing models. Future Internet. 2025; 17; 1: 28. DOI: 10.3390/fi17010028.

27. Vnedreniye standartov sovmestimosti uchebnykh instrumentov (LTI) = Implementation of the Learning Tools Interoperability (LTI) Standard [Internet]. Available from: <https://arcadia.spb.ru/studies/lti-for-elearning>. (In Russ.)

28. Moiseyeva N.A., Sotnikov A.D. Comparative analysis of neural network architectures for fake news detection. XV Mezhdunarodnaya molodezhnaya nauchno-prakticheskaya konferentsiya s elementami nauchnoy shkoly «Prikladnaya matematika i fundamental'naya informatika» = XV International Youth Scientific and Practical Conference with Elements of the Scientific School «Applied Mathematics and Fundamental Informatics» (Omsk, May 19–24, 2025). Omsk: Omsk State Technical University; 2025: 114–115. (In Russ.)

Сведения об авторах

Наталья Александровна Моисеева,
К.п.н, доцент, доцент кафедры «Прикладная математика и фундаментальная информатика» Омский государственный технический университет, Омск, Россия

Арсений Дмитриевич Сотников,
Разработчик
Seven Beats,
Омск, Россия

Information about the author

Natalia A. Moiseeva,
Cand. Sci. (Pedagogical), Associate Professor,
Associate Professor of the Department of Applied Mathematics and Fundamental Computer Science
Omsk State Technical University,
Omsk, Russia

Arseniy D. Sotnikov,
Developer
Seven Beats,
Omsk, Russia