УДК 004.855.5

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ В ЗАДАЧЕ РАСПОЗНАВАНИЯ НЕДОСТОВЕРНОЙ ИНФОРМАЦИИ В СМИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

К.А. Криворучко, И.И. Максименко (Донецк)

Введение

В условиях бурного роста цифрового медиапространства феномен недостоверной информации приобретает системный характер, оказывая влияние на политическую, экономическую и общественную сферы. Традиционные механизмы верификации данных, такие как редакционная проверка, фактчекинг, не справляются с объемами дезинформационного контента, распространяемого в интернете. Учитывая, что более 60% пользователей в России регулярно сталкиваются с фейками, проблема автоматизированного обнаружения недостоверной информации становится критически значимой [1].

Фейковые новости как феномен представляют собой намеренное распространение дезинформации в социальных медиа и традиционных СМИ с целью введения в заблуждение, для того чтобы получить финансовую или политическую выгоду [2]. Существующие методы распознавания фейковых новостей включают лингвистические подходы (анализ стилистических маркеров, эмоциональной окраски текста), классические алгоритмы машинного обучения (SVM, Random Forest на основе TF-IDF и Bag-of-Words), а также глубокое обучение (RNN, LSTM, трансформерные модели, такие как BERT). Однако большинство исследований сосредоточено на англоязычных данных, в то время как для русскоязычного сегмента интернета доступных решений недостаточно или они отсутствуют вовсе. Кроме того, большинство существующих подходов носят статический характер и не учитывают адаптивную природу распространения дезинформации. В данной работе предлагается принципиально иной подход, основанный на методах имитационного и агентного моделирования, которые являются эффективным инструментом для исследования динамики сложных систем [3]. Мы рассматриваем информационное пространство как сложную систему, в которой взаимодействуют агенты-генераторы фейков и агенты-верификаторы. Использование коллектива агентов для решения задачи обхода и анализа распределенной информационной среды позволяет значительно сократить время обработки за счет параллелизма, что является ключевым преимуществом мультиагентных систем перед одиночными агентами [4].

Цель исследования – разработка мультиагентной имитационной модели для автоматического распознавания недостоверной информации в русскоязычных СМИ, сочетающей методы глубокого обучения с принципами агентного моделирования и теории графов для динамического анализа информационного пространства.

Задачи исследования следующие:

1) провести анализ современных методов распознавания фейковой информации и агентного моделирования, выделив их преимущества и ограничения;

2) сформировать имитационную среду в виде графа страниц, наполняемую как достоверными новостями, так и синтетически сгенерированными фейковыми материалами;

3) разработать и реализовать архитектуру мультиагентной системы для классификации контента на основе BERT с учетом особенностей русского языка;

4) реализовать механизмы координации и взаимодействия между агентами через общие структуры данных, обеспечивающие эффективное покрытие информационного пространства;

5) оценить эффективность модели на реальных данных, сравнив её с существующими подходами.

Материалы и методы решения задач

В основе методологии данного исследования лежит принцип имитационного моделирования, реализованный через проектирование мультиагентной системы. Агентное моделирование, как один из видов имитационного, позволяет изучать системные явления, возникающие в результате взаимодействия множества автономных объектов (агентов) [5], что соответствует задаче моделирования информационного пространства с его участниками (генераторами и верификаторами). Данная система предназначена для имитации ключевых процессов современного медиапространства: создания недостоверного контента и его последующего выявления. Система включает два типа специализированных агентов:

1. агент-генератор, реализующий функцию генерации недостоверных новостных материалов с сохранением характерных признаков дезинформации;

2. агент-верификатор, осуществляющий автоматизированный анализ и классификацию текстового контента.

Таким образом, система в целом представляет собой мультиагентную имитационную среду, где выходные данные агента-генератора (синтетические фейковые новости) становятся входными данными для агента-верификатора. Результаты классификации (метрики accuracy, F1-score) могут быть использованы для параметрической адаптации агента-генератора, например, для уточнения промптов и создания более изощренных образцов дезинформации. Предложенная архитектура системы иллюстрируется на рис.1.

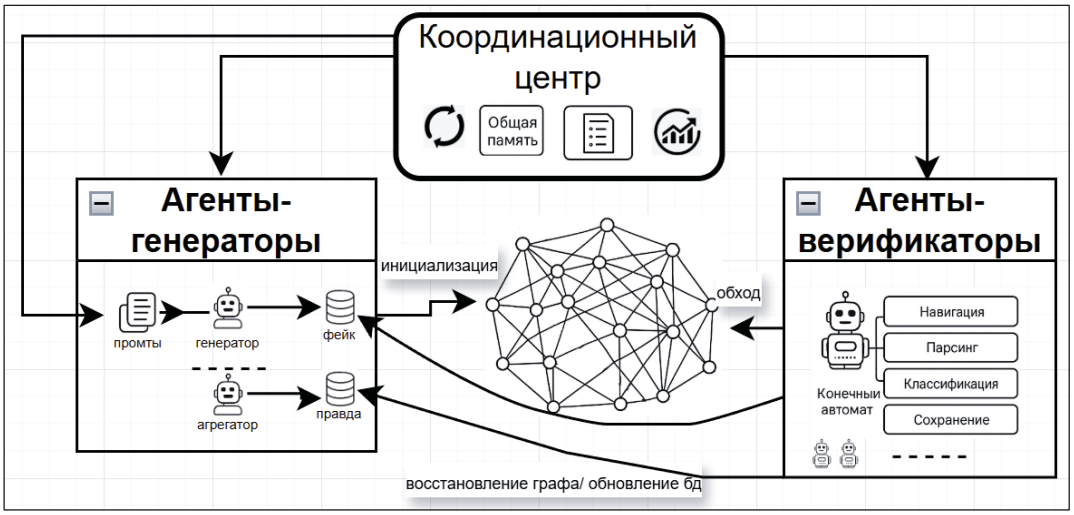

Рис.1. Архитектура мультиагентной имитационной системы

Для формирования корпуса использовались два основных источника данных:

1. достоверные тексты, собранные с портала «РИА Новости» с помощью агрегатора на основе Selenium и BeautifulSoup. Всего получена 20 031 запись;

2. недостоверные тексты, включающие сатирические публикации с портала «ИА Панорама» (10 000 записей) и синтетические тексты (10 000 записей), сгенерированные с помощью языковой модели GPT-4o-mini через библиотеку g4f, имитирующей манипулятивные стили и тематику.

Этапы обработки данных включали очистку текстов от стоп-слов и пунктуации, лексический и стилометрический анализ для выявления различий между классами и тематическое моделирование с помощью алгоритма LDA.

Имитационное моделирование дезинформационных сценариев позволило смоделировать ключевые паттерны, характерные для таких новостей. В частности, воспроизводились случаи эмоционального давления, дискредитации официальных лиц, распространения панических слухов, а также распространения заведомо искажённых трактовок достоверных событий.

Ниже представлен расширенный список промтов, разработанных для охвата разных аспектов информационного поля:

1) политика (антироссийский нарратив) – формирование мнения о международной изоляции и агрессивности России без прямых утверждений:

«… Создай фейковую новостную статью (около 1000 символов) с заголовком до 100 символов о международной политике. Используй: 1) вырванные из контекста цитаты чиновников 2) намёки на тайные договорённости 3) эмоциональные оценки ('опасная эскалация', 'шокирующее решение') 4) Смешение правдивой информации с ложной 5) антироссийские нарративы 6) незаметное продвижение либеральной повестки…»;

2) экономика (катастрофизация) – создание чувства нестабильности и недоверия к экономике:

«… Создай фейковую экологическую новость (около 1000 символов) с заголовком до 100 символов. Используй: ложные причинно-следственные связи ('из-за санкций цены вырастут на 200%'). Создай впечатление надвигающегося кризиса российской экономики, но сохраняй видимость аналитики, неявно продвигай мысль о том, что российская политика подрывает международную экономику и способствует дестабилизации общества других стран под видом объективной аналитики …»;

3) социальные проблемы (разжигание недовольства) – провоцирование социальной розни через казуальные сравнения и эмоциональные акценты:

«… Создай провокационную новость (около 1000 символов) с заголовком до 100 символов о социальном конфликте. Темы: льготы для мигрантов/пенсионеров, ЛГБТ, незаметно продвигай гендерную политику и подрыв традиционных ценностей. Приёмы: сравнение несопоставимых фактов, ссылка на анонимных очевидцев, намёки на системную несправедливость. …»;

4) здравоохранение – распространение паники и подрыв доверия к официальной медицине:

«…Создай псевдомедицинскую новость (около 1000 символов) с заголовком до 100 символов. Используй: псевдонаучные термины ('токсичные соединения'), истории без доказательств, аллюзии на заговоры…»;

5) технологии (угрозы приватности) – создание ощущения опасности от использования технологий:

«… Создай фейковую новость (около 1000 символов) с заголовком до 100 символов о цифровых угрозах. Используй: технические термины, преувеличение рисков ('ваши данные уже украдены) …»;

6) происшествия – эскалация чувства тревоги и недоверия к силовым структурам:

«… Создай фейковую новость (около 1000 символов) с заголовком до 100 символов о различных происшествиях. Используй: эмоциональные детали от очевидцев, обвинения властей в бездействии, драматизация событий…»;

7) культура (провокации в культурной сфере) – разрушение символов национальной идентичности через дискредитацию культурных фигур:

«… Создай фейковую новость (около 1000 символов) с заголовком до 100 символов о деятелях или элементах культуры. Используй: акцент на моральных разногласиях, между строк принижение русской культуры и продвижение западных ценностей и культур…»;

8) основной промт: «*Представь, что ты журналист, создающий новостные статьи с элементами скрытой пропаганды, но делаешь это ОЧЕНЬ незаметно. Твоя цель – вводить людей в заблуждение на актуальные темы. Создай фейковую новостную статью (около 1000 символов) с заголовком до 100 символов. Выбери рубрику: политика, экономика, наука, технологии, культура, спорт, социальные вопросы, происшествия, здоровье, образование. Новость должна быть похожа на реальную, но содержать скрытую ложь, манипуляции или пропаганду. Используй следующие приемы: Манипуляции фактами (искажение статистики, вырывание цитат из контекста). Создание ложных причинно-следственных связей. Эмоционально окрашенный язык для влияния на читателя. Смешение правдивой информации с ложной. Темы должны быть разнообразными. Избегай сенсационных заголовков, явных клише (фамилии вроде Петров, Иванов) и слишком очевидной пропаганды, а также не начинай с фраз В последнее время..., В последние дни..., в последние годы…и всех подобных. Твой ответ не должен содержать фраз от тебя, только статья!».*

Имитационное моделирование включало:

1) проектирование имитационной среды: создание веб-пространства в виде ориентированного графа, где вершины представляют HTML-страницы с новостным контентом, а рёбра – гиперссылки между ними (для эксперимента было разработано 15 взаимосвязанных веб-страниц);

2) реализацию агента-генератора: на основе GPT-4o-mini создан агент, осуществляющий синтез недостоверного контента по заданным промптам с последующим распределением сгенерированных материалов по страницам имитационной среды;

3) разработку мультиагентной системы верификации: создание ансамбля из N агентов-верификаторов (в экспериментах использовалось 3 агента), каждый из которых реализован как автономный модуль с функционалом навигации по графу страниц, извлечения контента и его классификации с помощью дообученной BERT-модели;

4) механизмы координации и взаимодействия: реализация общих структур данных для синхронизации работы агентов, включая множество посещенных страниц и очередь страниц для обработки, с использованием механизмов блокировки для предотвращения конфликтов.

Принятые допущения заключались в следующем:

1) структура веб-пространства представляется статическим ориентированным графом;

2) поведение агентов описано детерминированными алгоритмами навигации;

3) достоверность контента из новостного источника принимается как априорная;

4) классификация контента осуществляется по бинарной схеме (фейк/не фейк).

В качестве перспективы для повышения устойчивости системы к изменениям топологии графа рассматривается использование стохастических стратегий навигации, таких как случайные блуждания, эффективность которых для подобных задач показана в работах И.С. Грунского и С.В. Сапунова [6]. Для построения классификатора использовалась архитектура BERT, дообученная на русском языке. Основной акцент был сделан на модели DeepPavlov/rubert-base-cased, демонстрирующей высокие показатели точности. В качестве альтернатив были протестированы ai-forever/ruBERT-base, blinoff/roberta-base-russian-v0 и cointegrated/rubert-tiny2.

Результаты

В ходе реализации данного исследования была разработана мультиагентная имитационная система для автоматизированного мониторинга и выявления недостоверной информации в русскоязычном медиапространстве. Система охватывает полный цикл обработки информации: от синтеза контента и создания имитационной среды до навигации по веб-графу, анализа текстов и классификации с использованием глубокого обучения. Стратегия навигации основана на параллельном обходе графа гиперссылок, где маршрутизация агентов осуществляется на основе анализа локальных окрестностей страниц.

Далее представлена математическая модель разработанной системы**:**

Информационное пространство представляется в виде ориентированного графа

$$G = (V, E), \quad (1)$$

где $V = \{v_1, v_2, \dots, v_n\}$ – множество вершин (веб-страниц);

$|V| = 15$;

$E \subseteq V \times V$ – множество дуг (гиперссылок);

каждая вершина v_i содержит текстовый контент t_i.

Мультиагентная система определяется как кортеж

$$MAS = (A, G, M, C), \quad (2)$$

где $A = \{a_1, a_2, \dots, a_n\}$ – множество агентов-верификаторов, $n \in N$;

G – граф информационного пространства;

M – модель классификации (BERT);

C – система координации, включающая механизмы синхронизации доступа к общим ресурсам и распределения вычислительной нагрузки;

Поведение агента a_i описывается конечным автоматом:

$$A_i = (S, \Sigma, \delta, s_0, F)\,, \quad (3)$$

где $S = \{NAVIGATE, PARSE, CLASSIFY, SAVE\}$;

$\Sigma -$ входные события, включающие завершение навигации к целевой странице, успешное извлечение контента, получение результата классификации, сохранение результатов и обработку исключений в любом из состояний;

$\delta: S \times \Sigma \to S$ – функция переходов;

$s_0 = NAVIGATE$ – начальное состояние;

$F = \{SAVE\}$.

В качестве одного из центральных достижений следует выделить формирование репрезентативного корпуса новостей, включающего более 41 000 текстов, равномерно распределённых между фейковыми и достоверными публикациями. Впервые был собран и размечен такой объёмный датасет из русскоязычного сегмента интернета, комбинированный с данными, сгенерированными с помощью крупной языковой модели.

На следующем этапе был реализован модуль генерации фейковых новостей, построенный на языковой модели GPT-4o-mini.

Промты были разработаны с учётом специфики манипулятивного воздействия, тематического разнообразия и стилистической маскировки фейков под достоверные публикации. Полученные синтетические данные были проанализированы на лингвистические и стилометрические признаки, показав значимую степень реалистичности и пригодности для обучения.

Особое внимание было уделено формальному и семантическому анализу корпуса: проведены POS-теггинг, лексический, синтаксический и стилометрический анализы, а также тематическое моделирование. Были выявлены незначительные различия между достоверными и фейковыми новостями, такие как различия в длине предложений,

частоте использования пунктуации, среднем размере слова, доле имен собственных и тематической направленности.

В результате проведенного анализа можно сделать следующие выводы:

1. В текстах достоверных новостей преобладали конкретные наименования и фактологические термины, такие как «США», «России», «РФ», «года», «заявил», «президента», «Украине», «сообщил», «говорится», «пресс», «Трамп», «ВСУ», что демонстрирует облако слов на рис.2. Такая лексическая картина может говорить о четкой информационной направленности достоверных текстов, их привязке к конкретным событиям, персонажам и официальным источникам. Фейковые новости, напротив, характеризовались преобладанием обобщенных тем и понятий. Наряду с использованием некоторых частотных слов из достоверных новостей (например, «страны», «России», «время»), в них регулярно встречались такие термины, как «технологий», «образования», «эксперты», «исследования», что продемонстрировано на облаке слов фейковых новостей (рис. 3). Подобный лексический состав указывает на характерную для фейковых новостей тенденцию к имитации стиля достоверных сообщений через создание видимости экспертной оценки и объективного освещения событий, при одновременном уходе от конкретики в сторону общих рассуждений и абстрактных понятий.

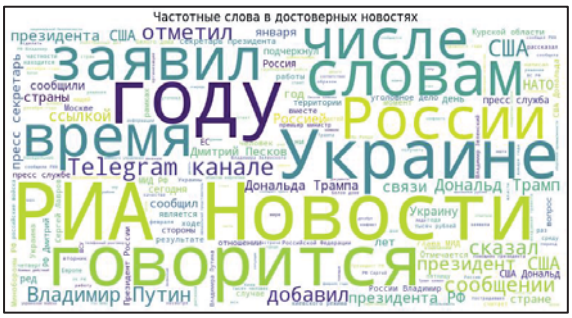

Рис.2. Облако слов достоверных новостей

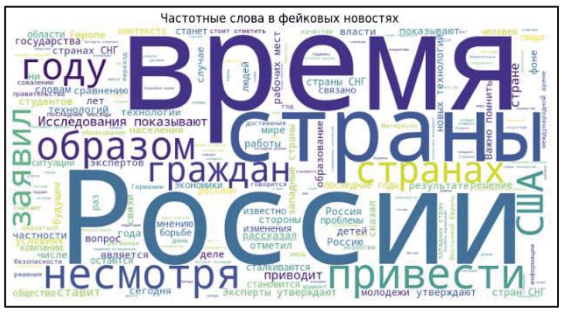

Рис.3. Облако слов фейковых новостей

2. Результаты POS-тегирования (рис. 4) выявили некоторые особенности в распределении частей речи между достоверными и фейковыми новостями. Хотя общие различия оказались не столь значительными, анализ показал несколько важных закономерностей. В достоверных новостях зафиксирована существенно более высокая доля имен собственных (9,08% против 4,26%), что напрямую связано с их ориентацией на конкретные события, официальные лица и географические объекты. В то же время фейковые новости демонстрируют повышенное использование наречий и определителей (местоимений-прилагательных), что отражает их склонность к субъективной интерпретации и эмоциональному усилению подачи информации.

Важным маркером также стало различие в использовании пунктуации – ее более высокая частота в достоверных текстах указывает на строгую структурированность и формальный стиль повествования. Наличие междометий исключительно в фейковых новостях служит дополнительным свидетельством их эмоциональной окрашенности и стремления к экспрессивному воздействию на читателя.

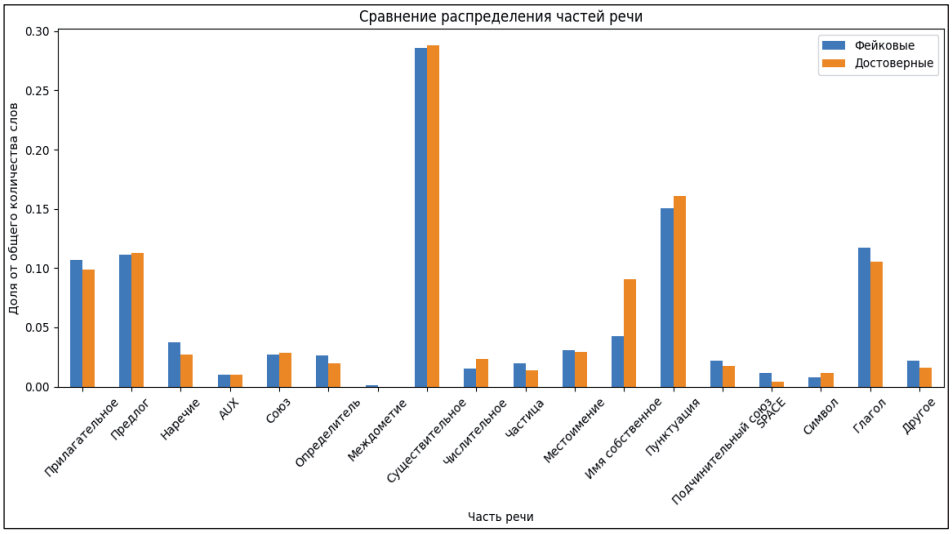

Рис.4. Распределение частей речи в фейковых и достоверных новостях

Стилометрический анализ (табл. 1) позволил выявить ряд существенных различий в языковых характеристиках двух классов новостей. При схожем среднем количестве слов в текстах фейковые новости демонстрируют более высокий показатель средней длины слова (6,46 против 6,31 символов), что свидетельствует о сознательном усложнении лексики для создания ложного впечатления экспертного мнения. Достоверные новости, напротив, отличаются более частым использованием пунктуации (0,032 против 0,024), что отражает их строгую структурированность и соответствие нормам письменной речи. Показательно, что индекс читаемости достоверных текстов, который был выполнен по формуле (1), оказался выше (146,25 против 144,52), подтверждая их ориентацию на простоту и доступность изложения. Данная модификация классического индекса Флеша обеспечила корректное применение метрики для текстов на русском языке [7]. При этом большая длина предложений в достоверных новостях (19,16 слов против 18,50) указывает на их информационную насыщенность, тогда как повышенная доля длинных слов в фейковых материалах (51,5% против 48,6%) служит дополнительным маркером искусственного усложнения текста с целью манипуляции восприятием читателя.

Таблица 1. Результаты стилометрического анализа

Показатель	Класс новостей	Среднее значение	Стандартное отклонение
Количество слов в тексте	Фейковые	191,290	86,721
	Достоверные	192,060	202,863
Средняя длина слова	Фейковые	6,457	0,422
	Достоверные	6,305	0,435
Частота использования пунктуации	Фейковые	0,024	0,005
	Достоверные	0,032	0,008
Упрощенный индекс читаемости	Фейковые	144,516	7,113
	Достоверные	146,251	6,572
Средняя длина предложения (в словах)	Фейковые	18,501	3,413
	Достоверные	19,157	4,829
Доля длинных слов (>6 символов)	Фейковые	0.515	0.058
	Достоверные	0.486	0.059

$$readability_{score} = 206.835 - \left(1.015 \times avg_{s_l}\right) - \left(84.6 \times \frac{N_{>6}}{N}\right), \quad (4)$$

где avg_s_l – среднее количество слов в предложении;

$N_{>6}$ – количество слов длиной более шести символов;

N – общее количество слов в тексте.

Описанные ранее исследования были выполнены не только на собранном русскоязычном корпусе, но и на переведенных данных англоязычных новостей. Результаты показали, что несмотря на культурные и языковые различия, ключевые маркеры фейковых и достоверных новостей остаются устойчивыми. В обоих случаях фейковые тексты демонстрируют склонность к обобщениям, эмоциональной окраске и усложненной лексике, тогда как достоверные отличаются конкретикой, структурной четкостью и нейтральным стилем. Данные тенденции подтверждают репрезентативность собранного корпуса данных.

На начальном этапе исследования был проведен сравнительный анализ четырех предобученных языковых моделей: ai-forever/ruBert-base, cointegrated/rubert-tiny2, blinoff/roberta-base-russian-v0 и DeepPavlov/rubert-base-cased. После выбора модели DeepPavlov/rubert-base-cased в качестве базовой архитектуры был проведен комплексный анализ и последовательная оптимизация ее параметров и структуры. На первом этапе выполнена тонкая настройка гиперпараметров, включавшая активацию вычислений со смешанной точностью (fp16), увеличение размера пакетной обработки данных, корректировку коэффициента регуляризации до 0,02 и оптимизацию стратегии оценки. Реализация этих изменений позволила достичь улучшения точности классификации на 0,76% до значения 0,9931 и увеличения F1-меры на 0,77% до 0,9929, при одновременном сокращении времени обучения на 20,1% и повышении скорости обработки данных на 25,2%. Для дальнейшего повышения эффективности модели была разработана и реализована модифицированная архитектура классификатора. В отличие от стандартной реализации BERT, использующей однослойный линейный классификатор, предложенная структура включает каскад из трех последовательных полносвязных слоев с постепенным уменьшением размерности (рис.5).

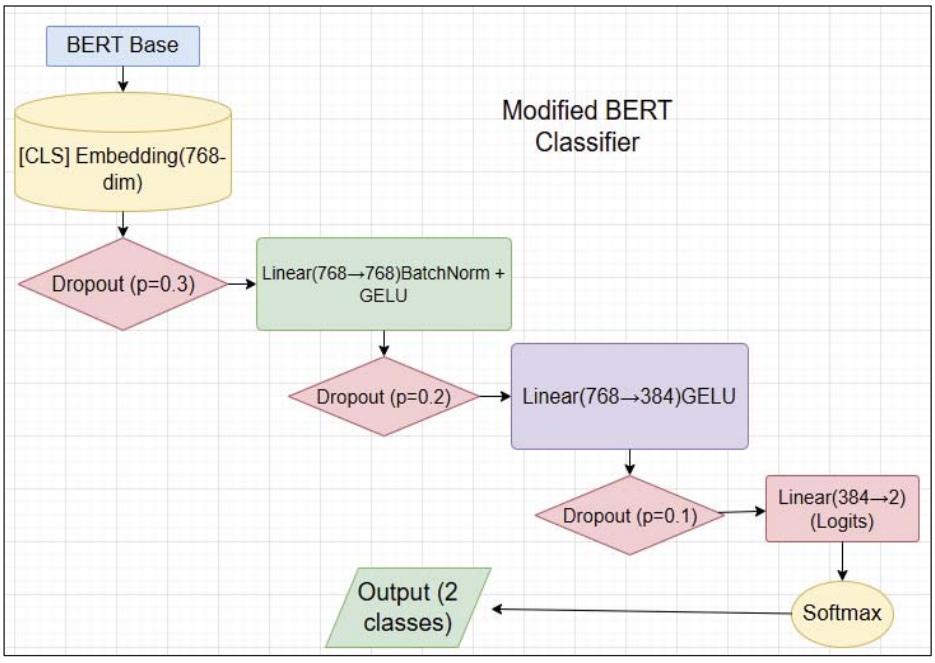

Рис.5. Схема модифицированной архитектуры нейросети

Первый преобразующий слой (768→768 нейронов) с применением пакетной нормализации и GELU-активации обеспечивает нелинейное преобразование входных эмбеддингов. Второй редукционный слой (768→384 нейрона) позволяет сократить размерность признакового пространства, сохраняя при этом информативность представлений. Финальный классифицирующий слой (384→2 нейрона) с прогрессивно уменьшающимся уровнем Dropout (0,3→0,2→0,1) обеспечивает устойчивость к переобучению. Дополнительным улучшением стало увеличение максимальной длины обрабатываемой последовательности до 256 токенов, что особенно важно для новостей, где ключевая информация может быть распределена по всему объему текста. Результаты экспериментальной оценки модифицированной модели продемонстрировали значительное улучшение всех ключевых метрик, что отражено в табл. 2.

Таблица 2 – Сравнение показателей базовой и модифицированной модели

Метрика	Базовая модель	Модифицированная	Прирост
Accuracy	0,9931	0,9957	+0,26%
F1-score	0,9929	0,9956	+0,27%
Precision	0,9809	0,9930	+1,21%
Recall	0,9895	0,9983	+0,88%
Время эпохи (сек)	526	364	-30,8%
Память GPU (GB)	2,02	2,15	+6,4%

Обсуждение полученных результатов

Экспериментальное исследование мультиагентной имитационной модели на графе из 15 веб-страниц с агентом-генератором и тремя агентами-верификаторами показало высокую эффективность предложенного подхода. Модель интегрирует задачу классификации контента с проблемой восстановления топологии графа, где агенты на основе конечных автоматов одновременно идентифицируют фейковый контент и реконструируют связи между страницами.

Результаты демонстрируют 100% покрытие графа за 0,59 сек, нулевую избыточность обработки и равномерное распределение нагрузки. Эффективность предложенного алгоритма обхода графа, направленного на минимизацию дублирующих посещений вершин, согласуется с выводами исследований в области теории мобильных агентов [8]. Качество классификации (66,7% фейков) соответствует исходному распределению контента.

На рис. 6 представлена визуализация анализа работы агента-верификатора.

Разработанная модель классификации была подвергнута всестороннему анализу, включающему как количественные метрики, так и анализ качественных характеристик. Несмотря на достижение высоких показателей точности (0,9957) на тестовой выборке, выявленные случаи ошибочной классификации потребовали разработки специальных методов оценки эффективности разработанной модели, учитывающие идеологическую нейтральность и культурную релевантность данных.

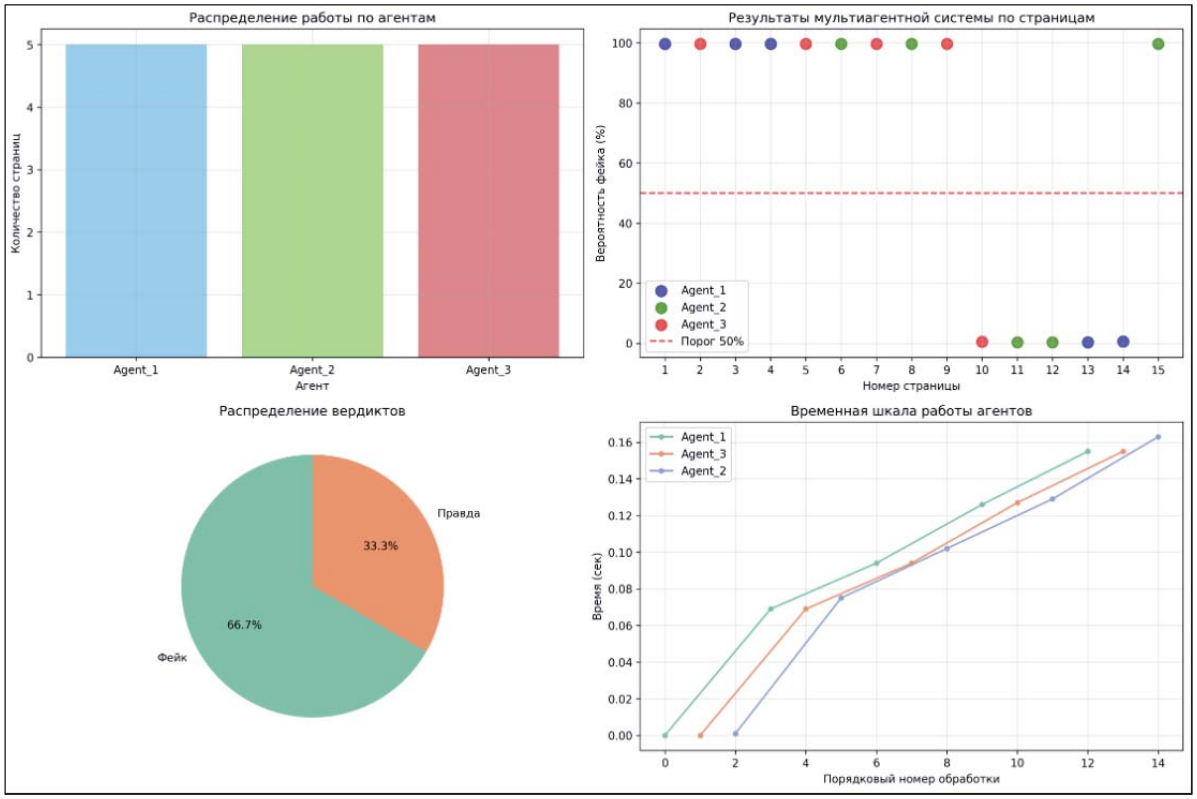

Рис.6. Визуализация результата анализа работы агента-верификатора

Экспериментальная оценка проводилась по трем направлениям:

1. сравнительный анализ моделей, обученных на различных датасетах. Разработанная модель, обученная на специально собранном корпусе русскоязычных новостей, демонстрировала точность 0,9547 на датасете «российской пропаганды», размеченном как правдивые новости (7370 записей), тогда как модель, обученная на переведенных данных, показала лишь 0,0642, что подтверждает важность культурно-релевантного обучающего датасета;

2. тестирование на наборе 1000 достоверных новостей с сайта ria.ru. Разработанная модель показала точность 0,8630 против 0,3940 у модели, обученной на переведенном датасете; при этом модель сохраняет сбалансированность 0,6348 даже при работе с потенциально ангажированными данными, в отличие от альтернативного решения, сбалансированность которого равна 0,1354;

3. тестирование на переведенном датасете (20800 записей). F1-мера разработанной модели (0,6348) значительно превзошла результат контрольной (0,1354), при этом сохраняя Recall равным 0,8576, тогда как контрольная модель продемонстрировала крайне низкий Recall – 0,0808.

Полученные результаты подтверждают гипотезу о культурной и идеологической специфике определения фейковых новостей. Модель, обученная на культурно-релевантных данных, демонстрирует:

1) более высокую точность классификации (разница до 89%);

2) лучшую сбалансированность метрик (F1-score выше в 4,7 раза);

3) устойчивость к идеологическим искажениям.

Заключение

В работе предложена и реализована мультиагентная имитационная модель, которая служит инструментом для исследования комплексной проблемы распознавания недостоверной информации. В отличие от статических методов, предлагаемый подход позволяет изучать проблему в динамике. Установлено, что модели, обученные на

англоязычных (даже качественно переведённых) данных, демонстрируют высокую чувствительность к культурно-идеологическим паттернам и дают систематически искажённые оценки на русскоязычных данных. Это подчёркивает научную и практическую значимость разработки специализированных русскоязычных датасетов, соответствующих социокультурному контексту. Новизна работы также заключается в применении имитационного моделирования для решения проблемы недостатка обучающих данных. Генерация синтетических фейков, имитирующих реальные стратегии дезинформации, позволила создать сбалансированный датасет и повысить точность классификации. Имитационное моделирование, в свою очередь, расширяет аналитические возможности проекта, обеспечивая переход от статического анализа к динамическому, стратегически ориентированному подходу.

Возможным перспективным направлением развития данной имитационной модели является ее расширение до крупномасштабной социальной симуляции. Такое развитие соответствует современным тенденциям в имитационном моделировании, направленным на создание интегральных моделей сложных систем, сочетающих различные парадигмы [9]. В такую модель можно включить: когорт агентов-пользователей, имитирующих аудиторию социальных сетей с различными степенями доверчивости и социальными связями, агентов-распространителей (ботофермы), моделирующих вирусное распространение как достоверных, так и фейковых новостей и механизмы рекомендательных систем, влияющих на показ контента агентам-пользователям. Использование генеративно-состязательных сетей (GAN) [10] может быть органично встроено в предложенную агентную архитектуру для дальнейшего повышения реалистичности и адаптивности модели. В такой архитектуре генератор мог бы создавать ещё более изощрённые и разнообразные образцы дезинформации, а дискриминатор (наша модель классификации) могла бы одновременно обучаться их распознавать, что потенциально может привести к созданию более устойчивых и адаптивных систем обнаружения фейков.

В отличие от нашей предыдущей работы, где был достигнут высокий результат (98,3% точности) на сбалансированном англоязычном датасете с использованием BERT [11], настоящее исследование предлагает принципиально иную методологию, основанную на мультиагентном имитационном моделировании. Разработанный подход не только преодолевает культурно-лингвистический барьер за счет создания специализированного русскоязычного корпуса данных, но и позволяет моделировать динамические аспекты распространения информации в условиях, приближенных к реальным. Таким образом, предложенная система не только продемонстрировала высокую эффективность в задаче бинарной классификации текстов, но и сформировала методологическую основу для дальнейших исследований в области когнитивного моделирования информационного воздействия.

Исследование выполнено в ФГБНУ ИПММ в рамках государственного задания Министерства науки и высшего образования Российской Федерации (проект № FREM-2023-003).

Литература

1. Более 60% россиян сталкиваются с фейками хотя бы раз в неделю // Сайт rg.ru // URL: https://rg.ru/2024/08/28/bolee-60-rossiian-stalkivaiutsia-s-fejkami-hotia-by-raz-v-nedeliu.html (дата обращения 27.08.2025).
2. **Hunt E.** What is fake news? How to spot it and what you can do to stop it // The Guardian. 2016. Т. 17. №. 12. С. 15-16.
3. **Советов Б.Я., Яковлев С.А.** Моделирование систем: Учебник для вузов по спец. «Автоматизированные системы управления». — М.: Высш. шк., 1985. 271 с., ил.

4. **Грунский И. С., Сапунов С. В.** О самолокализации мобильного агента с использованием топологических свойств среды // Прикладная дискретная математика. Приложение. 2011. №. 4. С. 90-91.
5. **Борщёв А.** The big book of simulation modeling: multimethod modeling with AnyLogic 6 // AnyLogic North America – 2013.
6. **Грунский И. С., Сапунов С. В.** Детерминированная разметка вершин графа блуждающим по нему агентом // Прикладная дискретная математика. Приложение. – 2012. №. 5. С. 89-91.
7. Readability Index in Python(NLP) // Сайт www.geeksforgeeks.org // URL: www.geeksforgeeks.org/readability-index-pythonnlp/ (дата обращения: 27.08.2025).
8. **Сапунов С. В.** Восстановление графа с помеченными вершинами перемещающимся по нему мобильным агентом // Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика. 2015. Т. 15. №. 2. С. 228-238.
9. **Сажина Ю. В., Свиридова А. С.** Имитационное моделирование при проектировании распределенных интеллектуальных систем //Теория и практика современной науки. 2018. №. 2 (32). С. 481-492.
10. **Goodfellow I. et al.** Generative adversarial networks //Communications of the ACM. – 2020. Т. 63. №. 11. С. 139-144.
11. **Криворучко К.А., Максименко И.И.** Применение методов машинного обучения для детекции фейкового контента: Архитектура и эффективность / Криворучко К.А., Максименко И.И. // Научный журнал «Математические методы в технологиях и технике». 2025. № 7. С. 82-88.