Леонова Юлия Сергеевна, инженер, Институт проблем управления имени В. А. Трапезникова РАН, г. Москва, Россия; e-mail: julia_leonova_123456@mail.ru, ORCID 0009-0008-7057-5244

Федянин Денис Николаевич, научный сотрудник, Институт проблем управления имени В. А. Трапезникова РАН, г. Москва, Россия; e-mail: dfedyanin@inbox.ru, ORCID 0000-0003-1032-5223

Чхартишвили Александр Гедеванович, доктор физико-математических наук, главный научный сотрудник, Институт проблем управления имени В. А. Трапезникова РАН, г. Москва, Россия; e-mail: sandro_ch@mail.ru, ORCID 0000-0002-2970-1244

РАСПОЗНАВАНИЕ ФАЛЬШИВЫХ НОВОСТЕЙ БОЛЬШИМИ ЯЗЫКОВЫМИ МОДЕЛЯМИ[1]

Статья основана на результатах исследования способности больших языковых моделей (LLM-моделей) различать достоверные и фальшивые новости. Специализированные организации способны затратить существенное количество ресурсов и провести доскональную проверку достоверности новости, однако читатели не всегда обладают столь мощными инструментами и оценивают степень достоверности исходя из своего опыта, мнения социального окружения и – в последнее время – результатов проверки новостей с помощью доступных читателям LLM-моделей. В результате исследования обнаружилось, что достоверные новости почти всегда верно идентифицируются LLM-моделями, а в идентификации фальшивых новостей ошибки встречаются довольно часто. Было уделено внимание и способности LLM-моделей корректировать фальшивые новости для придания им большей достоверности.

***Ключевые слова**: искусственный интеллект, большие языковые модели, фальшивые новости, проверка достоверности.*

Как цитировать: *Леонова Ю. С., Федянин Д. Н., Чхартишвили А. Г.* Распознавание фальшивых новостей большими языковыми моделями // Вестник Московского университета. Серия 10. Журналистика. 2025. № 5. С. 233–247. DOI: 10.55959/msu.vestnik.journ.5.2025.233247

Iuliia S. Leonova, Engineer, V. A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia; e-mail: julia_leonova_123456@mail.ru, ORCID 0009-0008-7057-5244

Denis N. Fedyanin, Research Scientist, V. A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia; e-mail: dfedyanin@inbox.ru, ORCID 0000-0003-1032-5223

Alexander G. Chkhartishvili, Doctor of Physical and Mathematical Sciences, Leading Research Scientist, V. A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Russia; e-mail: sandro_ch@mail.ru, ORCID 0000-0002-2970-1244

FAKE NEWS DETECTION BY LARGE LANGUAGE MODELS

The article is based on the results of a study evaluating the ability of large language models (LLMs) to distinguish between reliable and false news. While specialized fact-checking organizations are capable of conducting thorough investigations using substantial resources, ordinary readers typically lack access to such powerful tools. Instead, they assess the credibility of information based on personal experience, the opinions of their social environment, and increasingly, the output of publicly accessible LLMs. The study revealed that LLMs are highly accurate in identifying reliable news as such; however, they frequently make errors when classifying false news. The research also examined the capacity of LLMs to revise false news items in ways that make them appear more credible.

***Key words**: artificial intelligence, large language models, fake news, fact-checking.*

To cite this article: Leonova Iu. S., Fedyanin D. N., Chkhartishvili A. G. (2025) Raspoznavaniye fal'shivykh novostey bol'shimi yazykovymi modelyami [Fake news detection by large language models]. *Vestnik Moskovskogo Universiteta. Seriya 10. Zhurnalistika* 5: 233–247. DOI: 10.55959/msu.vestnik.journ.5.2025.233247

DOI: 10.55959/msu.vestnik.journ.5.2025.233247

Введение

Проблема широкого распространения фальшивых (фейковых) новостей в средствах массовой информации и социальных медиа в настоящее время приобрела глобальный характер. В условиях развития социальных медиа (в первую очередь социальных сетей и мессенджеров) эти новости могут распространяться молниеносно, в том числе через новостные сайты, воздействуя на общественное мнение. Особую остроту вопрос может приобретать в экономической сфере: шокирующие новости способны спровоцировать панику на биржах, повлиять на курсы валют и поведение инвесторов. Поэтому разработка эффективных методов выявления фальшивых новостей весьма актуальна, как актуальна и попытка привлечь для этого больше языковые модели (*large language models, LLMs*).

В работе K. Карамансьона (Caramancion, 2023) было проведено сравнение способности нескольких популярных *LLM*-моделей (*ChatGPT-3.5, ChatGPT-4.0, Google Bard* и *Microsoft Bing AI*) отличать реальные новости от фейковых. Наилучший результат показала модель *ChatGPT-4.0,* которая выявила 71% фальшивых новостей.

В работе китайских исследователей (Hu, Sheng, Cao, Shi et al., 2024) представлены результаты анализа других моделей (*Fine-tuned BERT, GPT-3.5, ARG, ARG-D*) на других датасетах. Усредненный результат этих моделей, измеренный метрикой *Macro F1*, достиг значения 0,784.

Отметим, что в англоязычном мире в качестве эмпирической базы для задач автоматического фактчекинга и детекции фейковых новостей широкое распространение получили три датасета: *PolitiFact, Gossip Cop* и *LUN*.

Датасет *PolitiFact* включает утверждения, проверенные политологами. Аннотированные классы охватывают диапазон от *True* до *Pants on Fire*, что делает этот ресурс важным для задач многоклассовой классификации достоверности. Кроме того, известны региональные (также англоязычные) разновидности, например *PolitiFact-Oslo* (Põldvere, Uddin, Thomas, 2023).

Датасет *Gossip Cop* ориентирован на новости о знаменитостях и развлекательном контенте, проверяемом на предмет вымышленных и спекулятивных утверждений. Оба корпуса интегрированы в состав *FakeNewsNet,* который обеспечивает единый формат данных и метаинформации (включая социальный контекст) для проведения комплексных исследований (Lin, Tremblay-Taylor, Mou et al., 2019).

Наконец, датасет *LUN* содержит тексты из трех категорий ненадежного контента — сатирического, пропагандистского и откровенно фейкового, а также надежные источники. Это позволяет использовать его в задачах бинарной классификации и изучения особенных признаков (например, стилистических различий между типами недостоверной информации) (Rashkin, Choi, Jang, Volkova et al., 2017).

Также известен аналогичный датасет *MOCHEG* (Yao, Shah, Sun, Cho et al., 2023).

При классификации датасетов трех различных моделей – *GPT-3.5, Instruct GPT, LLaMA2-13B* – было выявлено, что лучший результат показала модель *GPT-3.5* со значениями метрики *Accuracy* от 0,61 до 0,81 и метрики *F1* от 0,56 до 0,8 (Wu, Guo, Hooi, 2024).

Метрики качества для аналогичных датасетов трех моделей *GPT-4, GPT-3.5* и *Mistral7B* показали, что лучшей оказалась модель *GPT-4* (Vallidevi, Balaji, 2024).

Отметим, что выявление фейковых новостей возможно не только с помощью *LLM*, но и посредством обучения языковых моделей типа *BERT-MLP, BERT-CNN, BERT-LSTM* (см. Тумбинская, Галиев, 2023), но этот аспект выходит за рамки нашего рассмотрения. Целью данной работы является оценка и сравнение возможностей четырех современных *LLM* по обнаружению фальшивых новостей на материале экономических новостей на русском языке.

В некоторых работах подробно рассмотрено распознавание с помощью более простых моделей машинного обучения (Papageorgiou, Chronis, Varlamis, Himeur, 2024) и сетевые свойства фальшивых новостей (Choudhary, Arora, 2024).

Есть и отечественные исследования рассматриваемой в статье тематики (см., напр.: Третьяков, Филатова, Жук, Горлушкина и др., 2018; Бодулева, Сайфутдинова, Салина, 2020; Миронова, 2024; Усиков, 2024). В своих работах коллеги также изучают фальшивые новости, но их подходы отличаются от используемых в статье, так как затрагивают лишь общую роль фейковых новостей или посвящены использованию других методов определения фальшивых новостей. Например, в статье А. Р. Бодулева с соавторами (2020) описаны лингвистические методы определения таких новостей, но без использования современных средств машинного обучения и в качестве тестового взят другой набор данных.

Целью же данной работы является оценка и сравнение возможностей четырех современных *LLM* по обнаружению фальшивых новостей на материале экономических новостей на русском языке. Сначала мы кратко характеризуем исходные данные, на которых проверялись модели: новостные материалы из РБК как примеры заведомо достоверных сообщений и публикации ИА «Панорама» как заведомо фальшивые. Именно то, что новости совершенно точно фейковые, и обусловило выбор этого источника. После этого были вычислены метрики качества классификации различными *LLM* и проанализирована общая картина результатов.

Далее мы исследовали способность *LLM* преобразовывать фальшивые новости таким образом, чтобы они становились более похожи на достоверные. При этом критерием успешности такого преобразования считалось увеличение доли невыявленных фальшивых новостей.

Наконец, мы исследовали конфликтное взаимодействие (игру) двух акторов, первый из которых стремится генерировать правдоподобные фальшивые новости, а второй – распознать их.

Описание исходных данных и оценка способности *LLM* определять фальшивые новости

Для оценки способности *LLM*-моделей к различению достоверных и фейковых новостей были собраны два корпуса текстов – заведомо достоверные и заведомо фальшивые новости.

В качестве источника достоверных материалов использовался сайт информационного агентства РБК (*rbc.ru*), а в качестве источника фальшивых новостей – сайт сатирического проекта ИА «Панорама» (*panorama.pub*), публикующего тексты, стилизованные под реальные новости. С целью повышения тематической однородности публикаций брались заметки из разделов об экономике (*rbc.ru/economics* и *panorama.pub/economics* соответственно). Для текстового анализа использовались только заголовок и первый абзац (лид) заметки.

Анализ корпуса из 1000 отобранных публикаций РБК за период с октября 2024 по май 2025 г. (228 дней) показал, что в публикациях имеется выраженная политико-экономическая направленность с акцентом на международные отношения, в частности на взаимодействие России с другими странами, особенно с США. Лингвистический анализ датасета показал доминирование таких ключевых понятий, как «экономика», «власть», «нефть» и «газ», что отражает традиционную для РБК деловую направленность контента. Анализ *URL*-адресов показал равномерное распределение публикаций с небольшим преобладанием материалов за октябрь 2024 и апрель 2025 г. Общий характер контента соответствует деловой аналитике с элементами политического освещения, где ключевыми темами выступают государственное управление, международные отношения и экономические вопросы.

Анализ корпуса из 1000 публикаций ИА «Панорама» выявил выраженную экономико-политическую направленность издания с акцентом на внутреннюю повестку. В отличие от РБК, для которого характерен баланс между международной и российской тематикой, «Панорама» демонстрирует гораздо более яркую ориентацию на внутренние вопросы. Это убедительно подтверждается статистикой: упоминания «России» (332) встречаются в три раза чаще, чем

«США» (156). Отметим также, что особенностью заметок ИА «Панорама» является эклектичность тематического охвата – от бытовых вопросов до технологических трендов (биткоины, нейросети).

Для тестирования возможностей *LLM* были взяты четыре современные модели:

– *DeepSeek-R1* – это семейство открытых моделей рассуждения, демонстрирующих производительность, близкую к ведущим моделям, таким как *O3* и *Gemini 2.5 Pro*;

– *Qwen3* – новейшее поколение языковых моделей серии *Qwen,* включающее как плотные (*dense*), так и модели с использованием экспертов (*Mixture-of-Experts, MoE*);

– *Gemma3* – наиболее мощная на текущий момент модель, способная работать на одном *GPU*;

– *Mistral* – модель с 7 млрд параметров, выпущенная *Mistral AI,* обновленная до версии 0.3.

Участвующим в эксперименте *LLM*-моделям на вход подавался следующий промпт (текстовая инструкция): «Определи, является ли эта новость фейком или реальной Отвечай строго в формате: {{“verdict”: “real|fake”, “confidence”: 0.0–1.0}} Новость: [Новость]». *Confidence* (уверенность модели) – это вероятность, которую *LLM* вычисляет для своего решения (оценить новость как реальную (*real*) или фальшивую (*fake*)), отражающая степень уверенности в принятом решении.

Результаты, выданные *LLM*, обобщены в таблице 1.

Таблица 1

Эффективность LLM в определении достоверных и фальшивых новостей

LLM	Mistral	DeepSeek	Gemma	Qwen
Доля фальшивых новостей, правильно определенных как фальшивые	0,55	0,7	0,65	0,85
Доля фальшивых новостей, ошибочно определенных как достоверные	0,45	0,3	0,35	0,15
Доля достоверных новостей, правильно определенных как достоверные	1	1	1	0,95
Доля достоверных новостей, ошибочно определенных как фальшивые	0	0	0	0,05

Примеры не распознанных *LLM* фальшивых новостей:

«Союз заключенных России обратился к правительству России с требованием отменить новый сервис на портале Госуслуги, позволяющий пользователям установить самозапрет на заемные средства» (*Qwen 3*, *Gemma 3, DeepSeek*).

«Ленинский районный суд Магаданской области в своем недавнем решении признал найденные у гражданина наличные деньги признаком подготовки к совершению преступления. Судья объяснил, что "наличие у физического лица наличных денежных средств на руках не гарантирует их добросовестного использования и создает риски для общественной безопасности"» (*Mistral*, *DeepSeek*).

«Президент США Дональд Трамп на своей странице в соцсети *Truth Social* объявил о "фантастическом" плане по увеличению прибыльности американского сельского хозяйства. Американский лидер планирует заключить многомиллиардную "топинамбурную" сделку с Россией» (*Mistral*, *DeepSeek*).

«На портале «Госуслуги» появится раздел "Долг перед Родиной", где каждый россиянин сможет увидеть, сколько примерно он задолжал государству. При расчете будет учитываться, где живет гражданин, сколько ему лет, какими услугами он пользовался» (*Gemma 3*).

«В Южном Бутово открылся первый магазин новой федеральной сети дискаунтеров "Магнит вкуса", образованной после слияния популярных франшиз "Магнит" и "Азбука вкуса". Уже в день открытия жители столицы могли приобрести буханку хлеба всего за 800 рублей, а килограмм картофеля – по специальной акции за 999,99 рублей» (*Gemma 3, DeepSeek*).

«Следователям удалось установить основную причину блэкаута на Пиренейском полуострове. По мнению специалистов, виновница чрезвычайного происшествия – 55-летняя Долорес Ибаррури из Толедо» (*Gemma 3*).

«Российские банки получили право вводить комиссию за досрочное закрытие вкладов, которая может достигать 99% от суммы вклада. Это решение нацелено на стабилизацию банковской системы и защиту долгосрочных депозитных программ» (*Gemma 3*).

«Министерство труда и социальной защиты объявило о начале кампании по борьбе с финансовыми спекуляциями – граждане,

получающие доход от биржевых операций в размере, превышающем 100 тысяч рублей в год, будут привлекаться к ответственности как “лица, уклоняющиеся от общественно полезного труда”» (*Gemma 3*).

«Президент США Дональд Трамп заявил, что скоро победит в торговой войне с Китаем – он пообещал “продавить жуликоватого Винни-Пуха, ставшего на путь международного бандитизма”, намекая на председателя КНР» (*DeepSeek*).

Модель противоборства Фальсификатора и Разоблачителя фальшивых новостей

Чтобы предсказать выбор журналистами моделей ИИ для формирования фальшивых новостей и их разоблачения, была построена модель (игра), что позволило применить концепции равновесий (активно используемые в экономике для предсказания поведения рациональных субъектов) из математической теории игр. Необходимость применения этой теории заключается в следующем: оказывается, что не всегда лучшая в среднем модель разоблачения способна выявить лучшую в среднем модель генерации фальшивых новостей, и поэтому некорректно просто сравнить лучшую в среднем модель генерации и разоблачения. Ниже приведено краткое описание этой теории в объеме, необходимом для понимания полученных нами результатов.

Математическая теория игр – современная дисциплина, момент появления которой хорошо известен: в 1944 г. вышла монография Дж. фон Неймана и О. Моргенштерна «Теория игр и экономическое поведение» (1970), в которой были изложены известные на тот момент результаты.

Хотя и не сразу, но теоретико-игровое моделирование заняло прочное место в арсенале методов анализа социально-экономических процессов. Завершение этого процесса можно отнести, по-видимому, к 1990-м гг. В 1994 г. Нобелевскую премию по экономике получили Дж. Нэш, Дж. Харшаньи и Р. Зельтен, внесшие большой вклад в развитие теории игр, а в 1995 г. вышел учебник по микроэкономике (Mas-Colell, Whinston, Green, 1995), в котором имеется одноименный раздел (*Game Theory*). Из обширной литературы на русском языке отметим монографию о применении теории игр в управлении организационными

системами (Губко, 2002) и сравнительно недавнюю книгу (Захаров, 2015) с большим количеством примеров из социальных, экономических, политических, военных взаимодействий.

Взаимосвязь технологий искусственного интеллекта и теоретико-игрового моделирования в последние годы активно исследуется авторами как в целом (Han, Antonioni and Teixeira, 2025), так и в отношении конкретных предметных областей (см., напр.: Романова, 2025: 50–58; Serugga, 2025: 5, 39). В этом разделе мы рассмотрим простую теоретико-игровую модель (так называемую «матричную игру») взаимодействия двух акторов (игроков), использующих *LLM* для генерации и распознавания фальшивых новостей.

Первый актор (далее будем называть его «Фальсификатором») стремится сделать фальшивую новость правдоподобной. Точнее, он имеет дело с уже имеющейся фейковой новостью и стремится скорректировать ее таким образом, чтобы она стала как можно более правдоподобной. Предположим, что для этого он может использовать любую из рассмотренных выше *LLM (Mistral, Deepseek, Gemma, Qwen*).

Второй актор (далее будем называть его «Разоблачителем») стремится выявить фальшивую новость. При этом он также может использовать любую *LLM*.

Эмпирическое исследование показало, что средняя доля невыявленных фальшивых новостей (среди всех фальшивых новостей) различается для каждой *LLM* (см. табл. 2). Эти величины можно считать базовым уровнем, который Фальсификатор пытается повысить также при помощи *LLM*.

После того как оба игрока одновременно и независимо выбрали свои стратегии, Фальсификатор получит выигрыш, равный доле невыявленных (скорректированных *LLM* для придания им большей правдоподобности) фальшивых новостей. Естественно предположить, что выигрыш Разоблачителя противоположен выигрышу Фальсификатора.

На основании анализа приведенных выше эмпирических данных выигрыш Фальсификатора в зависимости от его стратегии и стратегии Разоблачителя представлен в таблице 2 (строки соответствуют стратегиям Фальсификатора, столбцы – стратегиям Разоблачителя).

Промпт для Фальсификатора приведен ниже: «Ты – журналист, который создает правдоподобные фейковые новости. Возьми эту

фейковую новость и измени ее так, чтобы она стала похожей на реальную. Сохрани стиль и тон оригинала. Верни только измененный текст в формате JSON с полями: {{“title”: “измененный заголовок”, “detailed”: “измененный текст новости”}} Исходная новость: [Исходная новость]».

Приведем пример использования работы Фальсификатора.

Оригинальная новость: «Ученые Европейского центра ядерных исследований (CERN) сообщили о прорыве в области квантовых технологий – с использованием последней версии квантового компьютера им удалось разработать устройство, способное генерировать биткоины».

Результат попытки повысить достоверность с помощью *Gemma 3*: «Генетический центр Европейского центра ядерных исследований (CERN) объявил о значимом достижении в области квантовых вычислений. Исследователи использовали передовую версию квантового компьютера для создания системы, способной производить и обрабатывать цифровые активы, в частности, криптовалюты. Хотя детали работы пока не разглашаются, представители CERN подчеркнули, что это открытие открывает новые возможности для исследований в области квантовых вычислений и потенциальных применений в финансовой сфере. Представители CERN заявили о проведении дальнейших экспериментов для оптимизации системы и изучения ее влияния на текущие финансовые рынки. Официальные комментарии по поводу масштаба и потенциального влияния этой разработки ожидаются в ближайшие недели».

Таблица 2

Доля невыявленных фальшивых новостей при стратегиях игроков

	Mistral	**Deepseek**	**Gemma**	**Qwen**
Mistral	0,35	0,35	0,55	0,1
Deepseek	0,9	0,8	0,8	0,4
Gemma	0,75	0,7	0,85	**0,45**
Qwen	0,65	0,55	0,75	0,25

Значения в таблице 2 полностью описывают модель матричной игры (антагонистической игры с двумя участниками и конечным множеством стратегий). Решением матричной игры является

ситуация равновесия – такая пара стратегий игроков, что каждому из них не выгодно ее менять. В данном случае, как легко убедиться, единственной равновесной ситуацией является пара стратегий (*Gemma, Qwen*), при которой доля невыявленных фальшивых новостей составляет 0,45 (в таблице 2 равновесное значение выделено жирным шрифтом).

Таким образом, по результатам анализа можно сделать следующий вывод: в игре Фальсификатора и Разоблачителя для первого оптимально (для придания фальшивым новостям видимости достоверных) использовать *LLM Gemma*, а для второго оптимально (для выявления фальшивых новостей) использовать *LLM Qwen*.

Выводы

На основании проведенных экспериментов с *LLM*-моделями можно сделать следующие выводы:

– достоверные новости почти всегда верно идентифицируются *LLM*-моделями как достоверные, а в идентификации фальшивых новостей ошибки встречаются довольно часто;

– из участвовавших в эксперименте *LLM*-моделей наиболее успешна в придании фальшивым новостям видимости настоящих модель *Gemma*, а для выявления фальшивых новостей – *Qwen*;

– найдены равновесные стратегии Фальсификатора и Разоблачителя.

Подводя итог, отметим, что, несмотря на впечатляющие успехи ИИ, решение задачи достоверного автоматизированного выявления фальшивых новостей является делом будущего.

Примечания

[1] Исследование выполнено при поддержке Российского научного фонда (проект № 25-18-00991) в части сбора данных для составления обзора раздела «Введение» и в части проведения эксперимента, описанного в разделе «Описание исходных данных и оценка способности LLM определять фальшивые новости», для сравнения возможностей различных LLM по обнаружению фальшивых новостей и исследования возможности LLM преобразовывать фальшивые новости для придания им большей достоверности.

Авторы благодарят Тимошкина Тимофея Александровича, студента ФГБОУ ВО «Российский биотехнологический университет», за выполненные им при прохождении практики в ИПУ РАН работы по формированию наборов данных.

Библиография

Бодулева А. Р., Сайфутдинова Л. Р., Салина А. С. Языковые особенности фейковых новостей (на материале фейковых новостей о Дональде Трампе и предвыборной кампании США 2020) // Вестн. Марийск. гос. ун-та. 2020. Т. 14. № 4 (40). С. 466–473.

Губко М. В., Новиков Д. А. Теория игр в управлении организационными системами. М.: Синтег, 2002.

Захаров А. В. Теория игр в общественных науках. М.: ИД Высшей школы экономики, 2015.

Миронова Н. Г. Технологии медиабезопасности: методы противодействия фейк-контенту в цифровых медиа // Экономика и право: проблемы, стратегия, мониторинг: монография. Чебоксары: ИД «Среда», 2024. С. 171–195.

Нейман Дж. фон, Моргенштерн О. Теория игр и экономическое поведение. М.: Наука, 1970.

Романова А. С. Теория игр для автономных систем искусственного интеллекта при управлении корпорациями // Надежность. 2025. Т. 25. № 2. С. 50–58.

Третьяков,А. О., Филатова О. Г., Жук Д. В., Горлушкина Н. Н. и др. Метод определения русскоязычных фейковых новостей с использованием элементов искусственного интеллекта // International Journal of Open Information Technologies. № 6. 2018. С. 99–105.

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.

Усиков Д. В. Приемы и методики распознавания контента, созданного с применением искусственного интеллекта // II Всероссийская школа НЦФМ по искусственному интеллекту и большим данным (25–29 ноября 2024 г.). Саров, 2024. С. 83.

Caramancion K. (2023) News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking. In *2023 IEEE Future Networks World Forum (FNWF)*. IEEE. Pp. 1–6.

Choudhary A., Arora A. (2024) GIN-FND: Leveraging users' preferences for graph isomorphic network driven fake news detection. *Multimedia Tools and Applications* 83 (22): 62061–62087.

Han T. A., Antonioni A., Teixeira A. S. (2025) Focus on game theory and AI in complex systems. *Journal of Physics: Complexity*. Available at: https://iopscience.iop.org/article/10.1088/2632-072X/ade588/pdf (accessed: 17.06.2025).

Hu B., Sheng Q., Cao J., Shi Y., Li Y., Wang D., Qi P. (2024) Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada. Vol. 38. Pp. 22105–22113.

Lin J., Tremblay-Taylor G., Mou G., You D., Lee K. (2019) Detecting fake news articles. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. Pp. 3021–3025.

Mas-Colell A., Whinston M. D., Green J. R. (1995) *Microeconomic theory*. New York: Oxford University Press.

Papageorgiou E., Chronis C., Varlamis I., Himeur Y. (2024) A Survey on the Use of Large Language Models (LLMs) in Fake News. *Future Internet* 16: 298.

Põldvere N., Uddin Z., Thomas A. (2023) The PolitiFact-Oslo Corpus: A new dataset for fake news analysis and detection. *Information* 14 (12): 627.

Rashkin H., Choi E., Jang J. Y., Volkova S., Choi Y. (2017) Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Pp. 2931–2937.

Serugga J. (2025) AI-Assisted Game Theory Approaches to Bid Pricing Under Uncertainty in Construction. *AppliedMath*. Available at: https://www.mdpi.com/2673-9909/5/2/39 (accessed: 17.06.2025).

Vallidevi K., Balaji V. (2024) Yours Truly: A Credibility Framework for Effortless LLM-Powered Fact Checking. *IEEE Access*.

Wu J., Guo J., Hooi B. (2024) Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Pp. 3367–3378.

Yao B., Shah A., Sun L., Cho J.-H., Huang L. (2023) End-to-End Multimodal Fact-Checking and Explanation Generation: A Challenging Dataset and Models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*. Association for Computing Machinery. Pp. 2733–2743.

References

Boduleva A. R., Sayfutdinova L. R., Salina A. S. (2020) Yazykovye osobennosti feykovykh novostey (na materiale feykovykh novostey o Donal'de Trampe i predvybornoy kampanii SShA 2020) [Linguistic features of fake news (based on fake news about Donald Trump and the 2020 US election campaign)]. *Vestnik Mariyskogo gosudarstvennogo universiteta* 14 (4): 466–473. (In Russian)

Caramancion K. (2023) News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking. In *2023 IEEE Future Networks World Forum (FNWF)*. IEEE. Pp. 1–6.

Choudhary A., Arora A. (2024) GIN-FND: Leveraging users' preferences for graph isomorphic network driven fake news detection. *Multimedia Tools and Applications* 83 (22): 62061–62087.

Gubko M. V., Novikov D. A. (2002) *Teoriya igr v upravlenii organizatsionnymi sistemami* [Game theory in the management of organizational systems]. Moscow: Sinteg Publ. (In Russian)

Han T. A., Antonioni A., Teixeira A. S. (2025) Focus on game theory and AI in complex systems. *Journal of Physics: Complexity*. Available at: https://iopscience.iop.org/article/10.1088/2632-072X/ade588/pdf (accessed: 17.06.2025).

Hu B., Sheng Q., Cao J., Shi Y., Li Y., Wang D., Qi P. (2024) Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada. Vol. 38. Pp. 22105–22113.

Lin J., Tremblay-Taylor G., Mou G., You D., Lee K. (2019) Detecting fake news articles. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE. Pp. 3021–3025.

Mas-Colell A., Whinston M. D., Green J. R. (1995) *Microeconomic theory*. New York: Oxford University Press.

Mironova N. G. (2024) Glava 14: "Tekhnologii mediabezopasnosti: metody protivodeystviya feyk-kontentu v tsifrovykh media" [Chapter 14: "Media security technologies: methods of counteracting fake content in digital media"]. In: *Ekonomika i pravo: problemy, strategiya, monitoring: monografiya* [Economy and law: problems, strategy, monitoring: monograph]. Cheboksary: Sreda Publ. (In Russian)

Neiman J. fon, Morgenstern O. (1970) *Teoriya igr i ekonomicheskoe povedenie* [Theory of games and economic behavior]. Moscow: Nauka Publ. (In Russian)

Papageorgiou E., Chronis C., Varlamis I., Himeur Y. (2024) A Survey on the Use of Large Language Models (LLMs) in Fake News. *Future Internet* 16: 298.

Põldvere N., Uddin Z., Thomas A. (2023) The PolitiFact-Oslo Corpus: A new dataset for fake news analysis and detection. *Information* 14 (12): 627.

Rashkin H., Choi E., Jang J. Y., Volkova S., Choi Y. (2017) Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Pp. 2931–2937.

Romanova A. S. (2025) Teoriya igr dlya avtonomnykh sistem iskusstvennogo intellekta pri upravlenii korporatsiyami [Game theory for autonomous AI systems in corporate management]. *Nadezhnost'* 25 (2): 50–58. (In Russian)

Serugga J. (2025) AI-Assisted Game Theory Approaches to Bid Pricing Under Uncertainty in Construction. *AppliedMath*. Available at: https://www.mdpi.com/2673-9909/5/2/39 (accessed: 17.06.2025).

Tretyakov A. O., Filatova O. G., Zhuk D. V., Gorlushkina N. N., Puchkovskaya A. A. (2018) Metod opredeleniya russkoyazychnykh feykovykh novostey s ispol'zovaniem elementov iskusstvennogo intellekta [Method for detecting Russian-language fake news using elements of artificial intelligence]. *International Journal of Open Information Technologies* 6: 99–105. (In Russian)

Tumbinskaya M. V., Galiev R. A. (2023) Identifikatsiya feyk-novostey s pomoshch'yu veb-resursa na osnove neyronnykh setey [Fake news identification using a neural network-based web resource]. *Programmnye produkty i sistemy* 4. Available at: https://cyberleninka.ru/article/n/identifikatsiya-feyk-novostey-s-

pomoschyu-veb-resursa-na-osnove-neyronnyh-setey (accessed: 17.06.2025). (In Russian)

Usikov D. V. (2024) Priemy i metodiki raspoznavaniya kontenta, sozdannogo s primeneniem iskusstvennogo intellekta [Techniques and methods for detecting AI-generated content]. In: *II Vserossiyskaya shkola NTsFM po iskusstvennomu intellektu i bol'shim dannym* [II All-Russian NTsFM School on Artificial Intelligence and Big Data]. Sarov. P. 83. (In Russian)

Vallidevi K., Balaji V. (2024) Yours Truly: A Credibility Framework for Effortless LLM-Powered Fact Checking. *IEEE Access*.

Wu J., Guo J., Hooi B. (2024) Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Pp. 3367–3378.

Yao B., Shah A., Sun L., Cho J.-H., Huang L. (2023) End-to-End Multimodal Fact-Checking and Explanation Generation: A Challenging Dataset and Models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*. Association for Computing Machinery. Pp. 2733–2743.

Zakharov A. V. (2015) *Teoriya igr v obshchestvennykh naukakh* [Game theory in the social sciences]. Moscow: Izdatel'skiy dom Vysshey shkoly ekonomiki Publ. (In Russian)

Поступила в редакцию
18.06.2025