

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ФАКТЧЕКИНГА: ТЕКУЩЕЕ СОСТОЯНИЕ, ВЫЗОВЫ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

И. А. Королёв

*Белорусский государственный университет,
ул. Кальварийская, 9, 220004, г. Минск, Республика Беларусь,
KorolevIA@bsu.by*

Рост количества циркулирующей в медиасфере информации, в том числе недостоверной, ставит перед журналистами новые вызовы. Одним из ответов на недостаток у редакций СМИ ресурсов по обработке больших объемов данных стало создание систем автоматизированного фактчекинга, работающих с применением технологий искусственного интеллекта. В статье на основании анализа научных работ, посвященных системам автоматизации фактчекинга, раскрыто современное состояние данного направления исследований, а также выявлены ограничения, с которыми автоматизированные системы сталкиваются на современном этапе своего развития.

Ключевые слова: фактчекинг; автоматизация фактчекинга; системы автоматизации фактчекинга; фейковые новости.

СІСТЭМЫ АЎТАМАТЫЗАВАНАГА ФАКТЧЭКІНГА: БЯГУЧЫ СТАН, ВЫКЛІКІ І ПЕРСПЕКТЫВЫ РАЗВІЦЦЯ

І. А. Каралёў

*Беларускі дзяржаўны ўніверсітэт,
вул. Кальварыйская, 9, 220004, г. Мінск, Рэспубліка Беларусь,
KorolevIA@bsu.by*

Рост колькасці інфармацыі, якая цыркулюе ў медыясферы, у тым ліку няпэўнай, ставіць перад журналістамі новыя выклікі. Адным з адказаў на недахоп у рэдакцый СМІ рэсурсаў па апрацоўцы вялікіх аб'ёмаў дадзеных стала стварэнне сістэм аўтаматызаванага фактчэкінгу, якія працуюць з выкарыстаннем тэхналогій штучнага інтэлекту. У артыкуле на аснове аналізу навуковых прац, прысвечаных сістэмам аўтаматызаванага фактчэкінгу, раскрыты сучасны стан дадзенага накірунку даследаванняў, а таксама выяўлены абмежаванні, з якімі аўтаматызаваныя сістэмы сутыкаюцца на сучасным этапе свайго развіцця.

Ключавыя словы: фактчэкінг; аўтаматызацыя фактчэкінгу; сістэмы аўтаматызацыі фактчэкінгу; фэйкавыя навіны.

Уже более десятилетия исследователи в области журналистики стремятся к созданию систем на основе искусственного интеллекта, позволяющих получать автоматический ответ на запрос о достоверности того или иного утверждения. В 2022 г. автором был проведен анализ функций некоторых из разработанных систем автоматизированного фактчекинга, на основании которого были выделены ограничения и перспективы развития данного направления [1].

За прошедшие с момента публикации статьи три года нейросети проделали исторический по своим масштабам и влиянию на медиасреду путь. При этом активное применение технологий искусственного интеллекта в значительной степени поспособствовало обострению проблемы фейковых новостей. Сегодня такие факторы, как рост количества дипфейков, постоянное совершенствование моделей генерации текста, все возрастающая сложность различения контента, созданного нейросетями, и контента, созданного человеком, повышают хаотичность медиасреды. Вдобавок к этому редакционный инструментарий верификации визуального контента, активно применявшийся журналистами еще несколько лет назад, требует кардинального обновления в эпоху технологий искусственного интеллекта. В частности, неэффективными становятся методы и инструменты анализа фотоконтента, позаимствованные журналистами из области судебной экспертизы и фотокриминалистики, такие как анализ уровня ошибок (ELA). Как пишет Н. А. Бондарева: «Поскольку нейросетевое изображение (если не подвергалось дополнительной обработке самим пользователем) не проходит череду преобразований и не подвергается сохранению с изменением форматов изображений, то методы, выявляющие артефакты при преобразовании и сжатии фотографий, с высокой степенью вероятности окажутся малоэффективными и полагаться на них нельзя» [2, с. 320].

В связи с этим целью данной статьи является актуализация информации о текущем состоянии систем автоматизированного фактчекинга, ревизия их эволюции и сохраняющихся ограничений.

На основании анализа академических работ ведущих исследователей в области автоматизированных систем фактчекинга, в частности, исследований Н. Диакопулоса, А. Ширама, К. Сильвермана, Н. Хассана, Б. Эдейра П. Накова И. Аугенштейна, Ч. Чхвэ, Э. Феррара и других, можно сделать следующие выводы о текущем состоянии развития данного исследовательского направления:

Во-первых, несмотря на определенный исследовательский интерес к системам автоматизированного фактчекинга, круг посвященной данной теме работ почти полностью ограничен исследованиями на английском языке и ни в коей мере не является сопоставимым по масштабам с исследовательским интересом к теме искусственного интеллекта и БЯМ (большим языковым моделям) в целом.

Во-вторых, современные большие языковые модели (БЯМ) эффективно расшифровывают аудиозаписи дебатов, интервью и новостных трансляций. Они также способны создавать текстовые документы на основе этих записей и выделять в них важные заявления [3]. Появляются системы, выполняющие узконаправленные функции, например, выявляющие новый контент социальных сетей, который либо подтверждает, либо противоречит утверждениям, ранее опровергнутым фактчекерами [4]. Однако, несмотря на успехи в выполнении отдельных частных задач, системы автоматизированного фактчекинга по-прежнему в полной мере не справляются с выполнением своей основной цели – проверки фактов.

В-третьих, до сих пор не появилось принципиального решения ни одного из упомянутых автором в статье 2022 г. ограничений автоматизированного фактчекинга, которые остаются все теми же: «недостаточный объем баз данных уже проведенных проверок фактов и необходимость их постоянной актуализации, охват исключительно англоязычного сегмента Сети, трудности в предоставлении аудитории контекста для корректной интерпретации проверенного факта» [1].

Учитывая вышеизложенное, можно констатировать, что за минувшие три года существенных сдвигов в решении ключевых проблем в работе автоматизированных систем фактчекинга не произошло. По-прежнему подобные системы могут решать лишь отдельные задачи, но полноценно служить комплексным инструментом в работе журналиста в данный момент они не способны. Другими словами, как и в начале своего развития, системы автоматизированного фактчекинга именно в области проверки утверждений (а не оценивания их принципиальной значимости и проверяемости) имеют функционал, сопоставимый с работой стандартных поисковых машин, что по-прежнему не позволяет им внедриться в практику работу редакций СМИ.

К тому же необходимо отметить, что проблема ограничений автоматизированных систем лежит не только в плоскости технологий, но затрагивает человеческий фактор, так как положенные в основу обучения

подобных систем базы данных не могут быть полностью объективны. Сложность минимизации человеческого фактора в системах автоматизированного фактчекинга состоит в том, что в отличие от поисковых систем, которые пусть и с ограничениями, но позволяют пользователю увидеть разные точки зрения по интересующему его вопросу, системы автоматизированного фактчекинга призваны дать один единственный ответ. И если даже этот ответ будет развернутым, очевидно, что он не может быть на 100% объективным.

Таким образом, необходимо констатировать, что с развитием технологий искусственного интеллекта разрыв между увеличением простоты, скорости и объемов создания разного рода фейкового контента и развитием систем автоматизации деятельности журналиста в области проверки фактов продолжает увеличиваться. В связи с этим резонно сконцентрировать больше внимания как исследователей журналистики, так и сотрудников редакций СМИ на иных способах противодействия фейковым новостям. В частности, можно предположить, что определенное позитивное влияние на упорядоченность медиасреды в аспекте феномена дипфейка могут оказать точечные законодательные меры, такие как маркировка контента, созданного с применением технологий искусственного интеллекта.

Библиографические ссылки

1. *Королёв И. А.* Автоматизация фактчекинга в журналистике: ограничения и перспективы развития // Труды БГТУ. Сер. 4, Принт и медиатехнологии. 2022. № 2(261). С. 50–56.
2. *Бондарева Н. А.* Графические нейронные сети и проблемы верификации изображений // «ГрафиКон»: материалы 33-ей международной конференции по компьютерной графике и машинному зрению, Москва, 19–21 сентября 2023 г.; Москва: Институт проблем управления им. В. А. Трапезникова Российской академии наук, 2023. С. 317–327.
3. Factuality challenges in the era of large language models / Augenstein I. [et al.] // Cornell University. 2023. URL: <https://arxiv.org/abs/2310.05189> (date of access: 08.06.2025).
4. *Choi E Ch., Ferrara E.* Automated Claim Matching with Large Language Models: Empowering Fact-Checkers in the Fight Against Misinformation // Cornell University. 2023. URL: <https://arxiv.org/abs/2310.09223> (date of access: 08.06.2025).