

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

ПРЕПРИНТ

ФЕЙК VS ФАКТЧЕКИНГ: ОТ АКАДЕМИЧЕСКИХ ДИСКУССИЙ К ПРАКТИЧЕСКИМ РЕШЕНИЯМ¹

Авторы:

Бережная Валентина Сергеевна, доцент
департамента медиа НИУ ВШЭ

Довбыш Ольга Сергеевна, доцент департамента
медиа НИУ ВШЭ

Ильвовский Дмитрий Алексеевич, доцент
департамента анализа данных и искусственного
интеллекта, научный сотрудник Международной
лаборатории интеллектуальных систем и
структурного анализа НИУ ВШЭ

Качкаева Анна Григорьевна, профессор
департамента медиа НИУ ВШЭ

Кирия Илья Вадимович, заместитель декана
факультета коммуникаций, медиа и дизайна,
профессор департамента медиа НИУ ВШЭ

Шомова Светлана Андреевна, профессор
департамента медиа НИУ ВШЭ

Москва 2022

¹ Препринт рабочего доклада подготовлен в рамках гранта исследовательского центра в сфере искусственного интеллекта «Межотраслевые технологии искусственного интеллекта и искусственный интеллект для иных приоритетных отраслей экономики и социальной сферы» АНО «Аналитического центра при Правительстве Российской Федерации» и Договора от 2 ноября 2021 года № 70-2021-00139 заключенного между АНО «Аналитический центр при Правительстве Российской Федерации» и ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики».

Аннотация. Категория фейка – одна из магистральных в современных исследованиях медиапроцессов, медиакультуры, медиаконтента и медиаграмотности. В современной комплексной медиасреде обычному потребителю крайне сложно ориентироваться, что упрощает попытки манипуляции контентом массовых информационных сервисов. «Фейковые новости», «постправда», разнообразные конспирологические теории и теории заговора и их распространение в цифровом пространстве становятся естественными в новой цифровой информационной парадигме. Согласно пресс-релизу ЮНЕСКО, самая опасная пандемия нашего времени – это «инфодемия», и данный термин даже вошел в язык Всемирной Организации Здравоохранения для обозначения информационной паники, сопровождающейся вспышками недостоверных новостей, слухов, теорий заговоров в электронных медиа.

В результате мы видим два направления, в которых может быть оказана помощь медиапотребителю:

- Развитие медиаграмотности (то есть способности самого потребителя критически мыслить, проверять факты, разбираться в источниках информации, уметь классифицировать источники и разбираться в степени правдоподобности фактов).
- Развитие автоматизированных информационных сервисов, которые помогают ориентироваться в большом наборе контента (например, рекомендательные сервисы), сортируют медиаресурсы по степени достоверности, помогают пользователям формировать нужные информационные предпочтения.

Большинство таких информационных сервисов имеют узкую корпоративную направленность или требуют специальных навыков для их использования, а также часто скрытым остается то, понимание фейка и фейковых новостей, которое было использовано в качестве теоретической базы при разработке алгоритмов распознавания, поэтому в настоящем исследовании мы обращаемся к исследованию базовых понятий.

В первой и второй части препринта представлен анализ и типологизация некоторых научных и практических подходов к определению фейковой информации (фейков) и фейковых новостей. В заключительной части приведены методики распознавания фейков, популярные инструменты фактчекинга, алгоритмы их работы и их ограничения.

Содержание

1 Fake (фейк): определение, дискуссии, типологические формы	4
2 Fake news: между публичной риторикой и научным направлением	7
3 Фактчекинг: платформы и особенности алгоритмов	15
Заключение.....	22
Приложение. Обзор проектов и инициатив по применению и развитию фактчекинга	23
Список использованной литературы	25

1 Fake (фейк): определение, дискуссии, типологические формы

К 2021 году в академической среде сложился целый спектр подходов к феномену fake (фейка) и тому, как определять fake news (фейк-ньюз). Авторы этих работ фокусируются на сложности для пользователей выявлять fake news онлайн, их распространении через тенденциозные медийные страницы, ответственности компаний социальных медиа и поисковых устройств по принятию мер против них и т.д.². Но главное, в чем сходятся самые разные исследователи, – это мнение, что регуляция распространения fakes и борьба с ними напрямую связана с консенсусом в определении самого понятия³, поэтому определение тезауруса представляется крайне важным.

Хотя и англоязычные, и русскоязычные теоретики пользуются разными обозначениями феномена, наиболее часто в литературе встречаются три понятия – misinformation (мисинформация – точного аналога этому понятию в русском языке нет), disinformation (дезинформация) и fake (фейк). Используется также термин hoax («утка») и некоторые иные, однако они значительно менее частотны, нежели первые три.

Как отмечают Дж. Фаркас и Я. Шоу⁴, «широко принятая и обсуждаемая терминология в рамках исследования ложной информации проводит различие между дезинформацией и мисинформацией. Хотя некоторые ученые используют эти термины как синонимы (Floridi 1996; Skinner, Martin 2000), большинство используют их, чтобы различать преднамеренные и непреднамеренные формы вводящей в заблуждение информации (misleading information). Некоторые ученые называют misinformation все виды информации, вводящей в заблуждение, а disinformation – только преднамеренное производство и распространение такой информации (Karlova, Fisher, 2013; Keshavarz 2014; Tadjman, Mikelic 2003). Другие используют термин misinformation в более узком смысле, чтобы охватить только непреднамеренные формы вводящего в заблуждение контента, что является противоположностью disinformation, которая включает только преднамеренные формы (Fallis 2015; Giglietto et al. 2016; Kumar, Wes, Leskovec 2016)»⁵.

Схожей позиции придерживаются и авторы гида университета Огайо (США), которые, ссылаясь на словарь «Dictionary.com», назвавший misinformation словом 2018 года, отмечают, что основная разница между mis- и dis- information заключается именно в «чистоте намерений» – в том, насколько автор осознает, что он распространяет не-истинную информацию. В случае с misinformation люди зачастую передают сообщение дальше, не зная, что оно не является истинным⁶. Disinformation – это намеренное введение других в заблуждение, и в этом принципиальная разница между явлениями.

Довольно часто fakes (фейки) характеризуются как явление, имеющее несколько типологических форм:

- сатира, «утка» (hoax), неточные новости (imprecise news), неадекватно переданные новости (poorly reported news), неправильное использование данных (misuse of data), неточная и небрежная журналистика (nonaccurate and sloppy journalism)⁷;

² Farkas J., Schou J. Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. *Javnost // The Public*. – 2018. – № 25(3). – P. 298-299.

³ Derakhshan H., Wardle C. *Information Disorder: Definitions, Understanding and Addressing the Disinformation Ecosystem*. Pennsylvania: Annenberg School for Communication. – 2017. – P. 5-12. – URL: <https://firstdraftnews.org/wp-content/uploads/2018/03/The-Disinformation-Ecosystem-20180207-v2.pdf> (дата обращения 26.04.2021).

⁴ См.: Farkas J., Schou J.

⁵ Там же, p. 299.

⁶ Мы определили бы это как «искреннее заблуждение автора».

⁷ Zaryan S. *Truth and Trust: How audiences are making sense of Fake News*. – MSc Media and Communication: Lund University – 2017.

- различные формы «вводящих в заблуждение новостей», приравняемых к фейкам: выделяются пропаганда, кликбейты, спонсируемые материалы, сатира и «утки», ошибки, тенденциозный контент, псевдонаука, мисинформация, фальшь⁸;
- не основанные на фактах или ложные новости (nonfactual / untrue news), слухи, конспирологические теории, сатира, фальшивые (неверные) утверждения политического руководства, необъективные или вводящие в заблуждение новости⁹.

В некоторых современных исследованиях фейки включаются в еще более широкий контекст – такой, например, как:

- «информационное загрязнение»¹⁰;
- медийные манипуляции (здесь, как правило, линия анализа восходит еще к работам У. Липпмана от 1922 года и доходит до самых современных работ, например, *Warwick, Lewis 2017*);
- информационная война¹¹ и др.

Фейки часто оказываются в общем ряду с пропагандой и конспирологическими теориями, лженаукой и обманными заверениями политиков, кликбейтом и слухами. Между тем многие авторы четко разводят, например, понятия фейковой информации и фольклорных проявлений недостоверного (или непроверяемого) контента. Так, *А. Гелферт*, давая определение фейкам, отличает их от таких категорий как сплетни, которые «релевантны только для специфических групп» и «распространяются очень избирательно в рамках установленных социальных сетей»; слухи, которые характеризуются более широкой аудиторией и представляют собой неавторизованные сообщения; «утки», которые, хотя и похожи на fake news, тем не менее создаются с целью в какой-то момент быть раскрытыми¹².

Близкую типологию предлагают и *Р. Гуадано и К. Гуттиери*¹³: они отличают фейки от слухов и фольклора, выстраивая следующий ряд родственных, но отличающихся друг от друга понятий: fake news – mis- или dis- information, замаскированная под новости, – слухи (широко циркулирующая неподтвержденная информация) – сплетни (опубличиваемая приватная информация) – городские легенды (страшные фольклорные истории) – заблуждения (вера, несмотря на факты, доказывающие обратное).

Особенно важно, что в зарубежных источниках наиболее общим является взгляд на фейк как на нечто, имеющее «злой умысел», то есть разновидность преднамеренной дезинформации. *А. Гелферт* определяет фейки как «намеренное представление (обычно) неверных или вводящих в заблуждение утверждений в виде новостей, где утверждения вводят в заблуждение умышленно»¹⁴. В этом смысле отечественная научная традиция порой придерживается противоположного мнения в определении фейка; в русскоязычной литературе встречается взгляд, который можно резюмировать как «фейк – от человека, дезинформация – от государства». *Г. Почепцов*, например, пишет: «Если мы разделим фейки и дезинформацию по их происхождению, то фейки являются результатом индивидуального производства, а дезинформация – институционального. В одном случае это человек, который может и ошибаться, поскольку в голове у него нет умысла обмануть

⁸ Steinberg L. Infographic: Beyond Fake News – 10 Types of Misleading News. – 2017. – URL: <https://eavi.eu/beyond-fake-news-10-types-misleading-info/> (дата обращения 26.04.2021).

⁹ Allcott H. (2017) Social Media and Fake News in the 2016 Election // *Journal of Economic Perspectives*. – 2017. – № 31(2). – P. 211-236.

¹⁰ Derakhshan H., Wardle C. Information Disorder: Definitions, Understanding and Addressing the Disinformation Ecosystem. Pennsylvania: Annenberg School for Communication. – 2017. – P. 5-12. – URL: <https://firstdraftnews.org/wp-content/uploads/2018/03/The-Disinformation-Ecosystem-20180207-v2.pdf> (дата обращения 26.04.2021).

¹¹ Khaldarova I., Pantti M. Fake News // *Journalism Practice*. – 2016. – № 10(7). – P. 891–901.

¹² Gelfert A. Fake News: A Definition // *Informal Logic*. – 2018. – № 38 (1). – p. 94.

¹³ Guadagno R.E., Guttieri K. Fake News and Information Warfare: An Examination of the Political and Psychological Processes From the Digital Sphere to the Real World // *Handbook of Research on Deception, Fake News, and Misinformation Online* / E. Chilwa, S.A. Samoilenko (eds.). – Hershey PA: IGI Global. – 2019. – p. 169.

¹⁴ См.: Gelfert A, p.85-86.

кого-то. В случае дезинформации исходным источником может быть структура, которая старательно скрывает свой характер. Ее действия направлены на то, чтобы ввести в заблуждение. Однако и те, и другие принадлежат к инструментам манипуляций»¹⁵.

Покажем это на некоторых примерах «коронавирусного дискурса». Например, широко разошедшееся весной 2020 года по социальным сетям шуточное стихотворение «Позвольте, жители страны...», приписываемое Пушкину (который якобы написал его в Болдино и «предугадал» нынешний карантин), на самом деле оказалось розыгрышем и может числиться, скорее, по ведомству сатиры или пародии, нежели по ведомству злостной дезинформации, намеренно вводящей людей в заблуждение. «Медицинские советы», которые распространяли люди не сведущие, но желающие помочь другим (принимайте от вируса имбирь, пейте чай с лимоном и т.д.) в рамках обозначенных выше подходов может именоваться скорее «мисинформацией» (искренним заблуждением), а не намеренным введением в заблуждение.

К. Уордл, одна из основательниц некоммерческого проекта First Draft News (их наглядная табличка с типами фейков получила большое распространение), объединяет мис- и дезинформацию в единое «пространство недостоверности», выделяя 7 общих для обеих категорий видов – от наиболее вредоносных, таких, как специально сфабрикованный контент или вводящий в заблуждение контекст, к наименее – сатире или пародии¹⁶.

Таким образом, можно резюмировать следующее:

Разброс определений фейка представляется достаточно широким. Среди научных дефиниций мы находим как те, которые смешивают fake и fake news и описывают последние как «новостные сюжеты», опирающиеся не на объективные или верифицируемые факты, доказательства, свидетельства и прочее, так и те, что радикально сужают феномен – например, называют фейком «вирусные посты фиктивных аккаунтов, которые созданы так, чтобы выглядеть как новостные сообщения».

До сих пор среди ученых не выработано единых параметров классификации фейковой информации и нет единства в понимании того, является ли фейк специфической категорией либо же это однопорядковое явление со слухами, городскими легендами и пр.

¹⁵ Почепцов Г. Фейки и дезинформация: в чем их сила? // Хвиля. – 2018. – URL: <https://hvylya.net/analytics/society/feyki-i-dezinformatsiya-v-chem-ih-sila.html> (дата обращения 26.04.2021).

¹⁶ Bakir V., McStay A. Fake news and the economy of emotions: problems, causes, solutions // Digital Journalism. – 2017. – № 6(2). – p. 1.

2 Fake news: между публичной риторикой и научным направлением

Помимо существования под «зонтичным брендом» фейка различных типов недостоверной информации, проблематичным нам видится и смешение в академическом тезаурусе понятий fake и fake news. Важно, что сегодня многие авторы используют эти термины как взаимозаменяемые.

Понятие fake news «фейковых новостей» в последние годы стало популярным в научных кругах, и, вероятно, в связи с популистским использованием данного термина за пределами академии. Предвыборная кампания кандидатов в президенты США Дональда Трампа и Хиллари Клинтон 2016 г. стала как раз спусковым крючком для проникновения этого понятия в популярный дискурс, а затем – и в академический. 8 декабря 2016 г. кандидат-демократ Хиллари Клинтон произнесла речь, в которой говорила об «эпидемии умышленного распространения фейковых новостей и фальшивой пропаганды, наводнивших социальные медиа последний год»¹⁷. Дональд Трамп даже учредил премию «фейковых новостей» для репортеров, допускающих, по его мнению, ошибки и неблагоприятные прогнозы, в частности тех, кто широко освещал так называемое вмешательство России в американские выборы. Так называемый «российский след» (англ. Russian Interference) после этого стал излюбленной темой западной прессы¹⁸. Динамика публикаций в академической базе научного цитирования Scopus ярко свидетельствует о том, что данный феномен попал в академическую среду akurat после выборов 2016 г. (Рис. 1).

Сам термин «фейковых новостей» в результате плотно вошел в различные академические сферы (от дискурсного анализа до компьютерных наук), однако его бытование продолжилось и в среде публичных интеллектуалов и политиков. При этом в этих кругах термин часто встречается вместе с другим термином – пост-правдой. В результате политики разных мастей к месту и не к месту стали использовать данный термин, некоторые даже предлагали «вакцинировать» право для борьбы с фейками¹⁹. Точно так же можно видеть, как сам термин «фейк» и его производные плотно вошли в обиход как политиков-популистов, клеймящих неправильно освещающие их деятельность медиа, так и системных политиков-законодателей, предлагающих все более жесткие и разветвленные методы для борьбы с фейками.

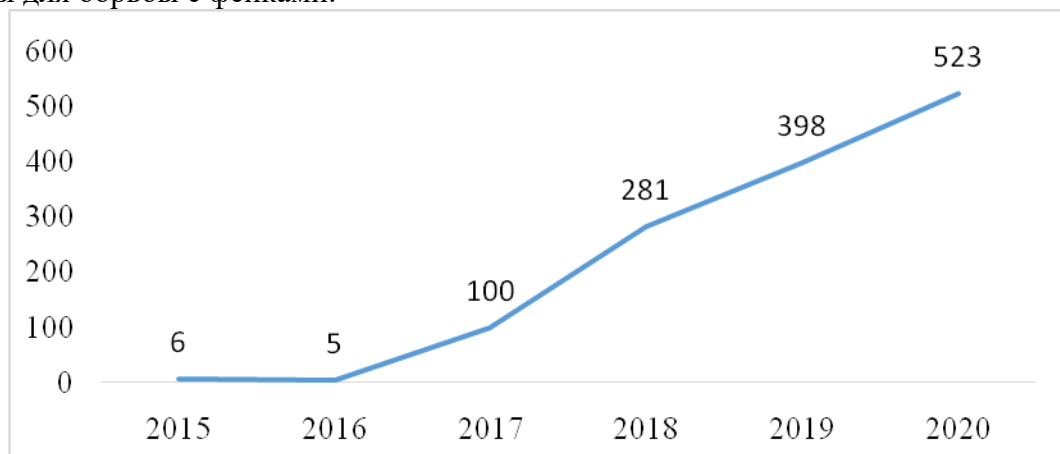


Рисунок 1 – Общее количество научных публикаций в базе Scopus со словосочетанием *fakenews*, отфильтрованные по области «социальные науки»

¹⁷ Wendling M. The (almost) complete history of “fake news” // BBC Trending. – 2018. – 22 January. – URL: www.bbc.com/news/blogs-trending-42724320 (дата доступа 01.12.2021).

¹⁸ См. напр.: Comai G. Responding to alleged Russian interference by focusing on the vulnerabilities that make it possible // Democracy and Fake News. Information Manipulation and Post-Truth Politics / ed. Giusti S., Piras E. – London: Routledge. – 2021. – P. 143-154.

¹⁹ Гордеев В. Бастрыкин предложил «вакцинировать» право для борьбы с фейками // РБК. – 2021. – 17 мая. – URL: <https://www.rbc.ru/politics/17/05/2021/60a271af9a79476afde3b2b6> (дата обращения 01.12.2021).

В России был введен закон о распространении фейковых новостей, в ЕС принят «План действий в области Европейской демократии» (European Democracy Action Plan), в котором сильно разросся и усилился блок, связанный с борьбой с дезинформацией²⁰. Причем важно понимать, что в основе как российских мер, так и европейского плана лежала идея фейковых новостей как «гибридных угроз» или средства ведения «гибридных войн». В этом, вероятно, и состоит «успех» концепта «фейковых новостей», что он используется одновременно для совершенно разных целей. Ян Красни справедливо замечает, что если изначально концепт фейков и пост-правды пришел из левых идей публичных интеллектуалов и профессоров-активистов, которые таким образом упрекали в предвзятости коммерческую неолиберальную прессу, то после выборов 2016 г. он превратился в «плавающее означаемое», которое стали использовать как левые и альтернативные политические силы в своей критике доминирующих элитных медиа, так и правые, а также либеральные корпоративные мейнстрим-медиа для упреков популистов, левых радикалов и т.п.²¹

При этом мы можем обнаружить несколько дискурсивных полей, в которых понятие «фейковых новостей» используется, что сильно затрудняет его концептуализацию. Вот ключевые из них:

- В контексте политической предвыборной борьбы в США со стороны правых популистов для дискредитации доминирующих коммерческих медиа (мейнстрим-медиа) и «демократических» политических сил;
- В контексте политической предвыборной борьбы в США со стороны либеральных демократических партий и их поддерживающих «мейнстрим-медиа» для дискредитации праворадикальных сил и популизма в лице президента Трампа.
- В контексте международной политики Западных стран для описания стратегии ряда противодействующих им стран (России, Китая) и их правительств. Часто это направление называют «гибридной войной». Сюда вписываются как речи американских демократов о вмешательстве России в американские выборы и распространение фейковых новостей в американском киберпространстве, так и Европейская риторика середины 2010-х годов, берущая курс на «борьбу со стратегической дезинформацией».
- В контексте международной политики России и некоторых других «незападных» стран для дискредитации информационной стратегии Западных стран, якобы направленной на сознательную эскалацию.

Такое «плавающее» состояние концепта «фейковых новостей» в публично-интеллектуальной, а также в научной среде говорит о том, что исследования в этой области пока лишь только набирают обороты и экспериментируют с научными методологиями. Однако, на что опираются в этих методологиях исследователи? Какие исследовательские направления и теоретическую базу они используют для этих целей? Для ответа на эти вопросы нами был произведен сетевой анализ и создание базы данных теоретических источников, на которые опираются данные работы.

Для начала в базе Scopus при помощи фильтров были выбраны только статьи, имеющие отношения к социальным наукам. Даже если часть из них относилась к компьютерным наукам, но индексировалась так же как «социальные науки», у нас были основания предполагать, что отнесение к категории социальных наук будет означать, что данные работы так или иначе будут рассматривать фейковую информацию как социальную проблему, а не как простой набор алгоритмов или базу данных, в которых можно

²⁰ European Commission Guidance on Strengthening the Code of Practice on Disinformation // Brussels: EC. – 2021. – URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/strengthening-eu-code-practice-disinformation_en (дата обращения 01.12.2021).

²¹ Krasni J. How to hijack a discourse? Reflections on the concepts of post-truth and fake news // Humanities and social sciences communications. – 2020. – № 7 (23). – P. 1-9. – URL: <https://doi.org/10.1057/s41599-020-0527-z> (дата доступа 26.04.2021).

валидировать/инвалидировать информацию на основе технических маркеров. Полученная база включала в себя 635 источников, и это была сама обширная область наук среди всех областей наук, где данное понятие встречалось (Рис. 2).

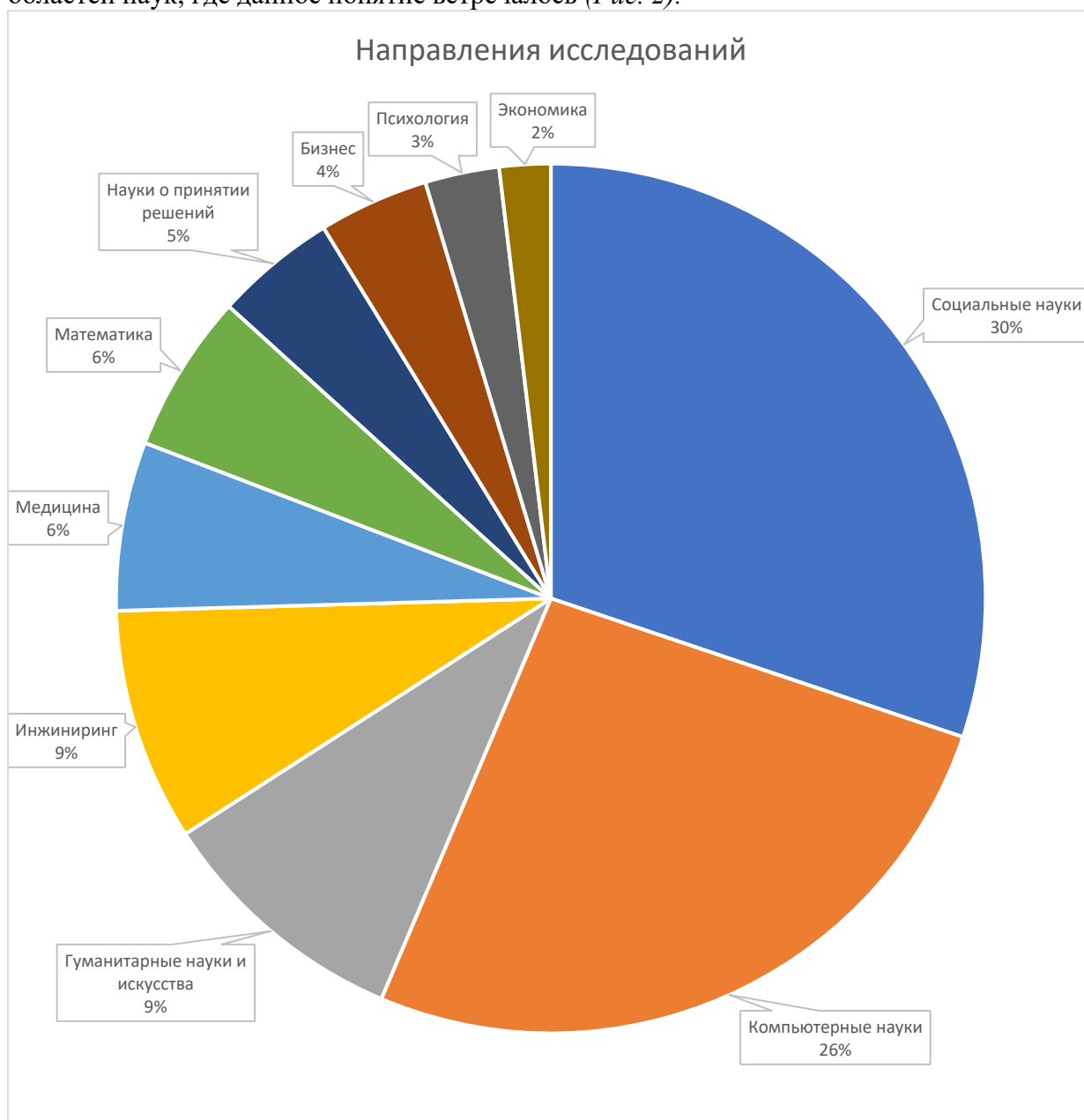


Рисунок 2 – Доля различных областей знаний в общем корпусе публикаций по категории «фейковые новости», $n = 1285$

На следующем этапе из списка по социальным наукам были взяты наиболее цитируемые 20 публикаций, то есть публикации, на которые чаще всего ссылаются другие авторы. Далее мы работали с библиографическими списками из этих 20 публикаций. Все источники были занесены в единую базу, из них был удален откровенный мусор, то есть источники, не имеющие никакого отношения к научным публикациям и, следовательно, не представляющие ценности как теоретическая база (аналитические отраслевые отчеты, обзоры, публикации СМИ и т.п.). Полученная выборка состояла из 542 источников. Каждый из источников был изучен, просмотрено его описание (абстракт) и собраны данные по его цитированию.

Мы условно разделили источники на три категории: классика (работы, изданные до 1990-го года), работы, изданные между 1990-2005 гг., и, наконец, все остальные. Такое

разделение для нас имело задачу разграничить классические теории медиа (до 1990 г.), работы, созданные в продолжение классических теорий, но по сути еще до появления партисипаторной культуры и социальных медиа, следовательно, так или иначе относящиеся к продолжению классики и, наконец, современные работы после 2005 г. опирающиеся как раз на более современную научную базу, связанную с новыми медиа. В результате на «классику» приходится 118 работ, тогда как на современный этап – 208 работ. Однако при этом, очевидно, наибольшее число совокупных цитат приходится именно на первую подгруппу (всего 202 567) (Рис. 3).

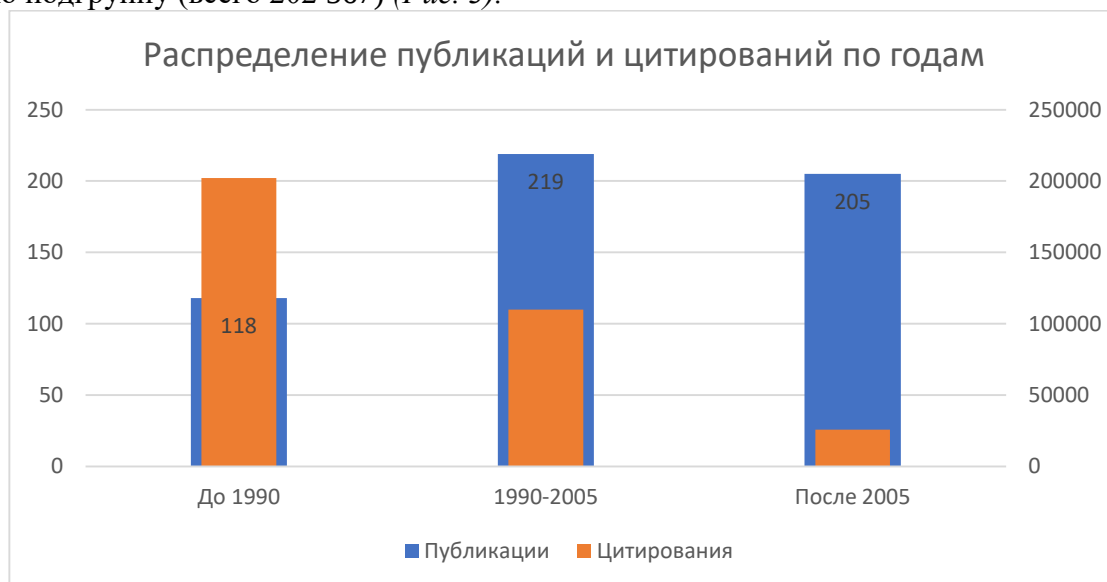


Рисунок 3 – Распределение публикаций и цитирований по временным отрезкам, $n = 542$

При этом на долю книг в массиве приходилось 230 публикаций (42%), на долю статей – 267 (49%). Остальную часть массива составляли главы из книг.

На следующем этапе исследования были построены сетевые графы, репрезентирующие отдельные работы и авторов, на которые ссылаются наиболее цитируемые работы по фейковым новостям. Итоговый граф изображен на Рис. 4.

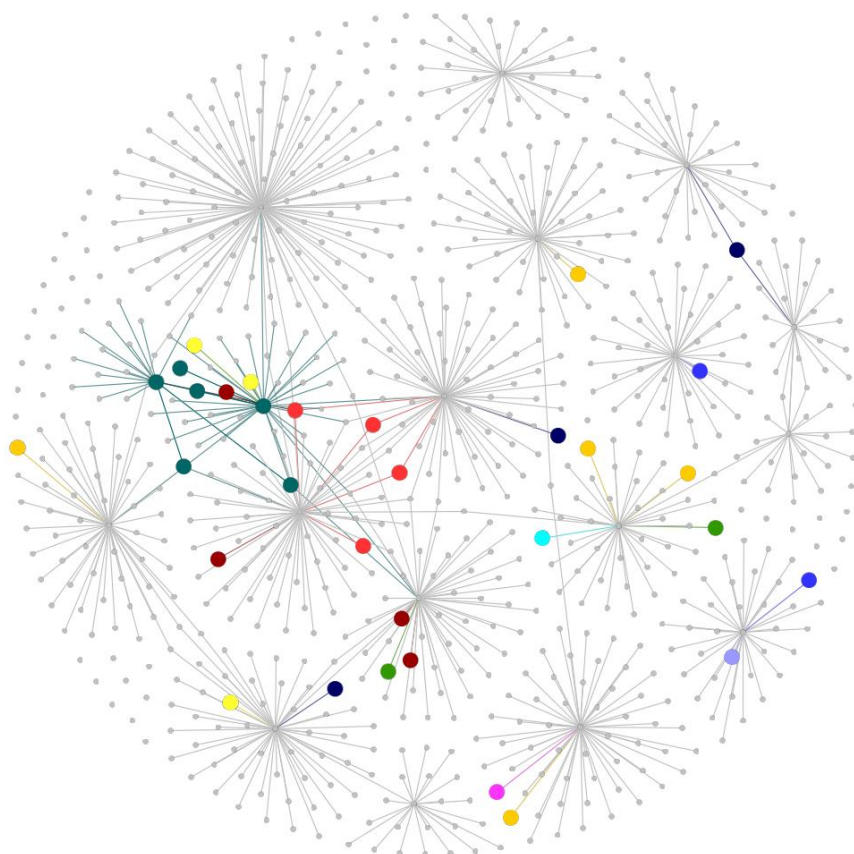


Рисунок 1 – Сетевой граф наиболее цитируемых работ и их источников

На этом графе видно, что в целом изученный массив публикаций содержит относительно небольшое количество кросс-цитируемых источников (источников, на которые ссылаются одновременно несколько публикаций). При этом существенный массив источников между публикациями не пересекается вовсе. То есть если в одной статье цитируется какая-то другая работа, маловероятно, что она же будет цитироваться в другой статье. Все это говорит о пока еще отсутствующем конвенциональном поле работ, которые являются общим базисом для исследований в области фейковых новостей. В центре нашего массива из 20 наиболее цитируемых публикаций можно видеть примерно 6-7 публикаций с достаточно плотной сетью взаимного цитирования и наличия общих ссылок.

Затем мы приступили к подробной работе с источниками. Перед исследовательской группой стояла задача – выявить теоретические группы, к которым относится каждый из источников (или подходы), а также выявить ключевые парадигмы исследований (более дробное деление теоретических подходов). В результате мы получили 14 теоретических подходов, к которым отнесены: Эмпирико-функционализм, Cultural studies, Медиасоциология, Медиапсихология, Философия медиа, компьютерный анализ медиа и т.п. Важно пояснить, что различие между эмпирико-функционализмом и медиасоциологией заключалось в том, что эмпирико-функционализм – классическая американская медиасоциология, ориентированная на изучение медиаэффектов и по сути своей давно устаревшая. Медиасоциология как правило применяет гораздо более широкий инструментарий и в некотором смысле ее можно назвать продолжательницей традиции эмпирико-функционалистских исследований. В схожих отношениях находятся, например, философия (к которому в том числе относится критическая теория, напр. Юрген Хабермас) и политическая экономия медиа, которая отчасти выросла из критической теории. Общий список в Табл. 1.

Media Sociology	114
Cultural studies	109
Empirical functionalism	88
Media Psychology	55
Linguistic	53
Political economy of media	31
Education science	25
Political science	25
Philosophy	19
Computer science	17
Librarian science	2
Micro-social theory of communication	2
Biology	1

Таблица 1 – Классификация источников по основным теоретическим подходам в массиве данных (кол-во ссылок)

Затем были выявлены ключевые научные парадигмы каждой из работ. Очевидно, набор парадигм был значительно длиннее, чем набор теорий, но именно набор парадигм позволяет в конечном счете судить о научном подходе и теоретических концептах, на которые опираются авторы в исследованиях. Самые часто встречающиеся парадигмы приведены в *Табл. 2*.

Парадигма	Количество источников
Медиаэффекты	105
Медиаграмотность	36
Репрезентация реальности в новостях	36
Популярная культура	35
Комедия/сатира	33
Дискурс	31
Рецепция (медиапотребление)	21
Менеджмент новостей	17
Публичная сфера	15
Контент-анализ	15

Таблица 2 – Классификация источников по ключевым научным парадигмам

Таким образом, можно увидеть, что наиболее популярными подходами, на которые опираются работы в области фейковых новостей, являются медиасоциология и эмпирико-функционализм, а также cultural studies. С нашей точки зрения, это демонстрирует два принципиально разных понимания феномена в академической среде.

На следующем этапе работы мы сократили итоговый массив данных и исключили из него статьи, оставив только книги. Исследовательская группа исходила из того, что в социальных науках самую высокую репутацию и теоретическую значимость имеют книги. Исследователи зачастую формулируют гипотезы и первые проекты каких-то теорий в статьях, а затем на основе этого выстраивается единая теория, которую автор воплощает в форме книги. Поэтому мы отбросили статьи и взяли именно книги. Сортировка книг по теоретическим подходам еще больше усилила позиции классических теорий медиаисследований: медиасоциологии, cultural studies, эмпирико-функционалистских работ, лингвистики, политической экономики медиа (*Табл. 3*).

Cultural studies	28
Media Sociology	24
Empirical fonctionnalism	16
Political economy of media	15
Linguistic	10

Таблица 3 – Группировка книг из списков источников по ключевым теоретическим направлениям

Для нас было важным понять также, насколько значимы и релевантны эти книги в целом в научном сообществе, поэтому мы отсортировали их по популярности/цитируемости их авторов. Полученные данные представлены на Рис. 5.

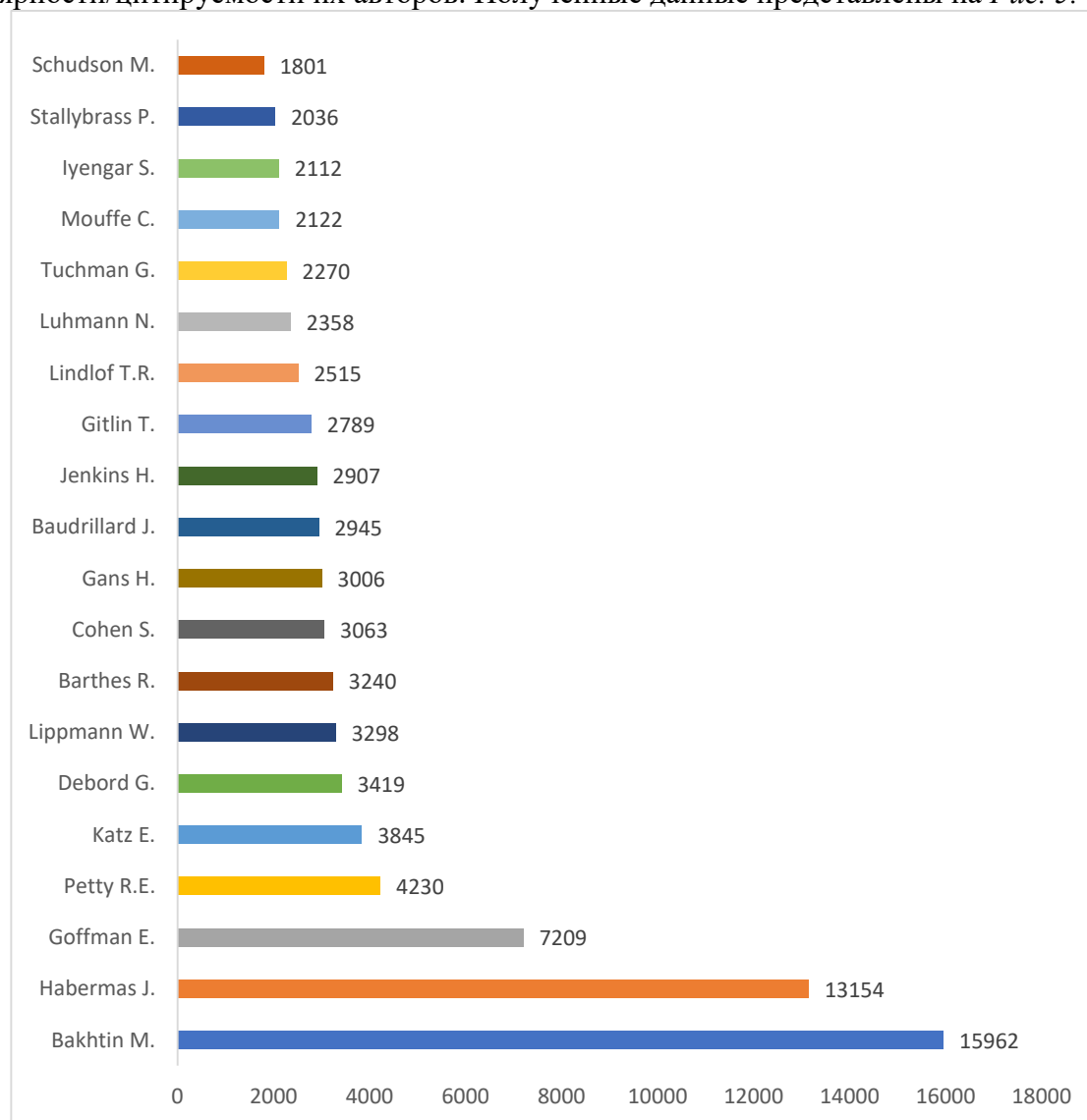


Рисунок 5 – Самые популярные (цитируемые) авторы из массива книг

Как следует из полученных данных, наиболее цитируемыми авторами, на которых ссылаются авторы работ по фейковым новостям, также относятся к двум главным выделенным категориям: cultural studies и исследования образов с одной стороны и медиасоциология – с другой. Здесь – классика медиасоциологии, медиаэффектов, исследований нарративов и образов и критической теории. Первую строчку занимает Михаил Бахтин, известный исследователь сатиры и карнавальной культуры. Затем следует Юрген Хабермас, представитель критической теории, далее – Ирвинг Гоффман и Ричард Петти – представители американской социологии коммуникаций.

Важно более подробно рассмотреть причины появления в топе списка исследователей семиотики и «конструируемой»/нереальной реальности (Барт, Дебор, Бодрийяр, Бахтин). Для этого мы обратились к более ранним работам по фейковым новостям и обратили внимание на существенное изменение данного понятия в академической среде после 2015 г., то есть после пенетрации понятия из публичной среды в академическую. До этого понятие «фейковые новости» обозначало телевизионные передачи пародийного жанра, посвященные высмеиванию актуальных событий (например, *The Daily Show*). В похожей трактовке фейковыми новостями будут считаться не «ложные новости», а заведомо ложные новости, которые никто изначально и не предполагал проверять на правдивость (например, различные пародийные интернет-сайты с фальшивыми смешными новостями и т.п.). Наиболее цитируемым автором здесь является Jeffrey Baym. В этом понимании термин начинает встречаться еще в 1990-х годах.

Актуальное понимание данного термина приходит после 2015, и именно здесь возникает понимание фейковых новостей как именно отклонения от базовых принципов новостей, в первую очередь принципа объективности и беспристрастности.

Таким образом, подводя итог, мы можем выделить две ключевых исторических традиции и два теоретических поля исследований фейковых новостей:

- Первое понимание связано с традицией, опирающейся на парадигму медиаэффектов, то есть воздействия масс-медиа на индивидов. Для этой традиции проблема фейковой информации – это проблема изменения политических эффектов масс-медиа, доверия к ним и политическим институтам и новостям. Работы, относящиеся к данной сфере, опираются на классику эмпирико-функционализма и отчасти медиапсихологии и цитируют работы Лазарсфельда, МакКомбса и Шоу и т.д. Для таких работ характерны алармистские настроения в отношении фейковых новостей, которые воспринимаются с функционалистской точки зрения как дисфункция масс-медиа, перверсивная форма масс-медиа и журналистики. Отчасти к этой же традиции относятся работы, анализирующие фейки в контексте нормативной традиции журналистики и рассматривающие этот феномен как отклонение от конвенциональной модели так называемой объективной журналистики.

- Однако второе понимание, которое мы выявили, оказалось несколько неожиданным и, очевидно, опирающееся на совсем иные представления о фейках. Это так называемые *cultural studies*, традиция популярной культуры, дискурсные исследования и исследования в сфере литературы и литературоведения (например, теория сатиры). Среди цитируемых авторов здесь можно встретить Хабермаса, Бахтина, Умберто Эко, Мишеля Фуко и т.п. С точки зрения этой традиции фейковые новости – это не перверсия, разрушающая общественную жизнь, а одна из культур восприятия реальности, наряду с другими. С этой точки зрения, никакой нормативной модели не существует, а существуют различные культуры восприятия реальности. Эта традиция отчасти опирается на работы постструктуралистов, считающих, что реальности не существует, так как любая реальность явлена нам посредством дискурсивных форм, которые сами по себе являются конструируемыми, то есть по идее могут быть другими. Фуко в связи с этим писал, что дискурсы конструируют конкурирующие друг с другом «режимы правды», между которыми происходит постоянная борьба.

3 Фактчекинг: платформы и особенности алгоритмов

Одним из наиболее логичных шагов в противодействии недостоверному контенту стали фактчекинг-сервисы и рекомендации различных интернет-ресурсов. Такие проекты, как «Медиа манипуляция» (mediamanipulation.org) с большим количеством разных кейсов и полезных определений или First Draft News (их миссия выражена на сайте вполне определенно: «Мы работаем, чтобы предоставить людям знания и инструменты для повышения устойчивости к вредоносной, ложной и вводящей в заблуждение информации»²²). Важную инициативу по объединению фактчекеров со всего мира предпринял Институт Пойнтера (poynter.org/ifcn) – их Международная сеть по проверке фактов не только устраивает конференции, отслеживает форматы и тенденции развития фейковых новостей, но и обеспечивает обучение онлайн.

Специальные советы, написанные доступным языком («Как говорить с детьми и подростками о дезинформации», «Семь типов дезинформации о выборах, на которые стоит обращать внимание» и т. д.), содержатся в блоге Массачусетского технологического института. В России известны такие проекты, как «Лапшеснималочная» А.Ковалева (в последнее время этот ресурс не обновляется) и «Проверено. Медиа» И. Бера, FakeCheck.ru М.Товкайло и Д.Казьмина. Немалую просветительскую роль играют и специальные «гиды» от различных университетов: например, подробный путеводитель по фейкам и недостоверной информации от Университета Огайо или гид Массачусетского технологического института по дипфейкам²³. Стоит обратить внимание при этом, что эти ресурсы отличаются различной степенью подробности, оригинальности, качества рекомендаций.

Поскольку социальные медиа стали одной из главных мишеней, по адресу которых летят обвинения в распространении фейков, – главы этих компаний не только приглашаются для отчетов в правительственные и регулирующие органы зарубежных стран, но и дискутируют с законодателями по важнейшим вопросам ответственности за контент, допустимости цензуры и т. д.²⁴, – вполне понятна заинтересованность крупных платформ в противодействии дезинформационным процессам. При этом в разных частях света мы видим самые разные юридические и этические кейсы как попыток саморегуляции платформ, так и взаимодействия их с органами власти.

В Америке и в Азии платформы микроблогов и социальные медиа также активно (хотя и по-разному с точки зрения смысловых векторов взаимодействия) участвуют в регулировании фейкового контента. Еще в 2018 г. китайская платформа Вейбо (Weibo) дала «заслуживающим доверия СМИ» и правительству Китая право напрямую пометать сообщения, распространяющие фейковые новости, специальным уведомлением, определяющим их как ложные или вводящие в заблуждение. «Уведомление появится сразу под сообщением, при этом Вейбо не играет никакой роли в его проверке или утверждении»²⁵.

Такие платформы как Twitter, Youtube, довольно давно политику ограничений распространения контента некоторых типов, один из которых – недостоверный. Так, в «Правилах Twitter» отдельный параграф посвящен подлинности распространяемой на

²² См.: firstdraftnews.org

²³ Somers M. Deepfakes, explained // MIT Sloan. – 2020. – URL: https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained?utm_source=mithp&utm_medium=web&utm_campaign=deepfakeexplainer (дата обращения 26.04.2021).

²⁴ Левченко Г. Главы Twitter, Google и Facebook приняли участие в онлайн-слушаниях в Сенате США, посвященных модерированию контента. Цукерберг не сразу дозволился // Медуза. – 2020. – URL: <https://meduza.io/feature/2020/10/28/glavy-twitter-google-i-facebook-prinyali-uchastie-v-onlayn-slushaniyah-v-senate-ssha-posvyaschennyh-moderirovaniyu-kontenta-tsukerberg-ne-srazu-dozvonilsya> (дата обращения 26.04.2021).

²⁵ Ziyu Y. Weibo Gives Media, Government Power to Quash «Rumors» // Sixth Tone. – 2018. – URL: <http://www.sixthtone.com/news/1003152/weibo-gives-media,-government-power-to-quash-rumors> (дата обращения 26.04.2021).

платформе информации, где особое внимание уделяется политике в отношении манипуляции платформой и спама (правила обновлены в сентябре 2020 г.)²⁶.

В России социальные медиа и государственные организации также озабочены усиливающимися процессами распространения нежелательного контента. Скажем, платформа рекомендательного сервиса Яндекс.Дзен осенью 2020 г. запустила фактчекинговый проект, привлекая к сотрудничеству не только теоретиков, но и индустриальных партнеров: ТАСС, Интерфакс, Ведомости, The Bell, Проверено.Медиа. Программа Я.Дзен по проверке информации работает второй год²⁷. В ходе проверки блогов платформы на наличие недостоверного контента, уже просмотрены тысячи публикаций. При этом фактчекеры далеко не всегда сходились во мнениях, оценивая ту или иную публикацию, и платформой были сформулированы важные для дальнейших дискуссий. Например, как проверять видео, которые могут оказаться дипфейками; оценивать ли публикацию как фейк, если автор излагает факты верно, но делает неверные выводы; как быть с информационными поводами, вокруг которых идет информационная война или муссируется большое количество слухов, и т. д. Здесь мы вновь возвращаемся к проблеме тезауруса.

Сервис Яндекс.Кью почти одновременно с Я.Дзеном запустил сообщество «Альянс русскоязычных фактчекеров», объединяющего профессиональных фактчекеров (директор ЦЦКМ А. Качкаева – член Альянса). «Альянс» в октябре 2021 году провел первую Школы «В поисках правды» для блогеров и заинтересованных журналистов, работающих на платформе Яндекса.

Сегодня все больше университетов и НКО занимаются проектами, которые помогают рядовым пользователям проверять информацию, их можно разделить на следующие категории:

1. Подборки наиболее широко распространенных сообщений, в отношении которых был проведен фактчекинг (преимущественно НКО).

2. Подборки полезных источников, позволяющих лучше понять, что такое фейк и как его распознавать (в основном, это делают университетские центры, иногда представляющие собой отдельные подразделения, но иногда такое можно найти и на сайтах крупных компаний).

3. Расширения для браузеров или приложения, которые можно установить и которые проведут проверку сами (в этих случаях довольно сложно судить о методологии, поскольку она раскрывается в очень сжатом виде, если вообще раскрывается).

Обзор некоторых проектов и инициатив приведен в Приложении 1.

Было предложено множество формулировок задач для решения проблемы распространения дезинформации в Интернете, и для каждой формулировки было исследовано множество подходов. Далее приведен обзор исследований фактчекинга и алгоритмов его работы.

Одним из релевантных исследований является работа по оценке достоверности информации в Twitter, представляющая подходы к задаче классификации тем из трендов на «заслуживающие» и «не заслуживающие» доверия (*Castillo et al., 2011*)²⁸. Для решения использовались признаки на основе поведения пользователей, текстов сообщений, тем и путей распространения информации.

²⁶ Политика в отношении манипуляции платформой и спама // Twitter. – 2020. – URL: <https://help.twitter.com/ru/rules-and-policies/platform-manipulation> (дата обращения 26.04.2021).

²⁷ Яндекс.Дзен. – URL: <https://yandex.ru/support/zen/requirements/fact-checking.html> (дата обращения 26.04.2021).

²⁸ Castillo C., Mendoza M., Poblete B. Information Credibility on Twitter // In Proceedings of the 20th International Conference on World Wide Web (WWW '11) / Association for Computing Machinery. – Hyderabad. – India. – 2011. – P. 675–684.

Rashkin et al. (2017)²⁹ проанализировали лингвистические особенности, которые используются в новостных сообщениях, содержащих сатиру, искажения или пропаганду. Для этого был собран корпус текстов с Politifact без мета-информации, дополненный новостными статьями из различных пропагандистских и не заслуживающих доверия источников.

В работе Wang (2017)³⁰ представлен схожий с Politifact набор данных LIAR для проверки фактов с использованием только проверяемого утверждения. В датасете каждое утверждение сопровождается соответствующей мета-информацией, содержащей сведения об авторе утверждения (аффилиация, текущее место работы). В качестве бейзлайна авторы предложили метод на основе лингвистических паттернов.

Lee et al. (2021)³¹ выявили, что дезинформация может быть обнаружена с помощью анализа значения перплексии языковой модели. В основе этого результата лежало наблюдение, что перплексия достигает более высоких значений для ложных утверждений. Также был проведен ряд исследований о возможности использования языковых моделей для генерации ответов на отдельные вопросы или для проверки фактов (Lee et al., 2020; Petroni et al., 2019; Roberts et al., 2020)³². Так, языковая модель должна была сгенерировать, например, дату или город в проверяемом факте. Однако исследования показали, что даже лучшие языковые модели сами генерируют недостоверные сведения.

Тексты с недостоверной информацией отличаются наличием специфических лексических и семантических средств, способствующих более быстрому распространению. Эти средства объединяются под словом «пропаганда». В 2019 году был представлен датасет для распознавания пропаганды в новостных статьях (Da San Martino et al., 2020)³³. Авторы сформулировали две подзадачи для sharedtask по данному датасету: (1) выделение пропагандистских фрагментов в тексте и (2) классификация каждого из них на одну или несколько из 18 пропагандистских техник. Наиболее успешными оказались подходы на основе языковых моделей с архитектурой Transformer (BERT, RoBERTa).

Объяснимый фактчекинг. Данное направление более актуально, чем фактчекинг не предполагающий доказательств, так как вместе с результатом проверки алгоритм предоставляет и его аргументацию. Например, в качестве аргументации может быть использован результат проверки одного из ранее рассмотренных утверждений. В этом случае задача состоит в поиске утверждения с тем же смыслом в корпусе проверенных ранее. Shaar et al. (2020)³⁴ разработали два набора данных для обнаружения ранее

²⁹ Rashkin H., Choi E., Jang Jin Yea, Volkova S., Choi Y. Truth of Varying Shades: Analyzing Language in Fake News and Political FactChecking // In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP '17). Association for Computational Linguistics. – Copenhagen. – Denmark. – 2017. – P. 2931–2937.

³⁰ Shu K., Sliva A., Wang S., Tang J., Liu H. Fake News Detection on Social Media: A Data Mining Perspective // SIGKDD Explor. Newsl. – 2017. – № 19, 1. – P. 22–36.

³¹ Lee N., Bang Y., Madotto A., Fung P. Towards Few-shot Fact-Checking via Perplexity // In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT '21). Association for Computational Linguistics. – Online. – 2021. – P. 1971–1981.

³² См. работы: Lee N., Li B., Wang S., Yih Wen-tau, Ma H., Khabisa M. Language Models as Fact Checkers? // In Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER '20). Association for Computational Linguistics. – Online. – 2020. – P. 36–41. Petroni F., Rocktäschel T., Riedel S., Lewis P., Bakhtin A., Wu Y., Miller A. Language Models as Knowledge Bases? // In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP '19). Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 2463–2473. Roberts A., Raffel C., Shazeer N. How Much Knowledge Can You Pack Into the Parameters of a Language Model? // In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP '20). Association for Computational Linguistics. – 2020. – Online. – P. 5418–5426.

³³ Shaar S., Babulkov N., Da San Martino G., Nakov P. That is a Known Lie: Detecting Previously Fact-Checked Claims // In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL '20). Association for Computational Linguistics. – Online. – 2020. – P. 3607–3618.

³⁴ Shaar S., Nikolov A., Babulkov N., Alam F., BarronCedeno A., Elsayed T., Hasanain M., Suwaileh R., Haouari F., Da San Martino G., Nakov P. Overview of CheckThat! 2020 English: Automatic Identification and Verification of

проверенных утверждений, которые были расширены и использованы для задач в рамках CLEFCheckThat! в 2020 и 2021 годах (Barrón-Cedeño et al., 2020; Hasanain et al., 2020; Nakov et al., 2021; Shaar et al., 2021).

Еще одним источником доказательств или опровержений могут являться графы знаний. Vlachos and Riedel (2015)³⁵ и Thorne and Vlachos (2017)³⁶ использовали для проверки численных утверждений subject-predicate-object (SPO) триплеты. Итоговый вердикт получался с помощью правил для сравнения значений из утверждения и извлеченного триплета. Gad-Elrab et al. (2019)³⁷ применили семантическую трассировку по текстам и графам знаний для реализации системы Tracy. В ней SPO факты объясняются или опровергаются с помощью логической агрегации информации, извлеченной из графов знаний. Так как не любую информацию можно представить в графе, подобный подход имеет ограничения. Поэтому Ciampaglia et al (2015)³⁸ предложили применять вероятностное прогнозирование на основе топологии графа.

Отдельно можно выделить проверку фактов на основе данных из Википедии. Thorne et al (2018) представили корпус FEVER, в котором для проверки утверждения требуется текстовый evidence (доказательств) из Википедии. Каждой проверяемое утверждение в паре с evidence размечено, как один из трех классов: support, refutes, notenoughinfo. Далее если извлеченное доказательство было неверным, то общий результат проверки признавался неверными вне зависимости от правильности итогового вердикта. Корпус FEVER отличается большим размером – в нем представлено 185.000 размеченных утверждений, что позволило исследователям обучать большие языковые модели. Набор данных использовался в задачах FEVER и FEVER 2.0 (Thorne et al., 2018b; Thorne et al., 2019), где большинство предложенных систем имело следующие компоненты: (1) поиск документов, (2) поиск релевантных предложений и (3) логический вывод на естественном языке (NLI). Для поиска релевантных документов применялись поисковые API (Chakrabarty et al., 2018; Hanselowski et al. 2018; Alonso-Reina et al., 2019). Для поиска релевантных предложений использовались методы на основе расстояния перемещения слов (Chakrabarty et al., 2018), TF.IDF (Malon, 2018), BERT (Stammbach and Neumann, 2019); а для определения логического следствия модели DAM (Otto, 2018), ESIM (Hidey and Diab, 2018), RandomForest (Reddy et al., 2018), LSTM (Nie et al., 2019) и BERT (Stammbach and Neumann, 2019).

Тем не менее важно отметить, что описанные выше методы проверки фактов на основе текстовых данных (текстовый evidence) не могут обеспечить полноту процесса верификации, так как многая информация представлена в виде таблиц и баз данных. Chen et al. (2020) предложили датасет TabFact для проверки фактов против структурированных данных, в котором 16.000 таблиц были размечены в качестве доказательств/опровержений для 118.000 утверждений. В качестве бейзлайна к решению были предложены подходы Table-BERT (кодирует таблицу и проверяемые утверждения в вектора для последующего сравнения) и LatentProgramAlgorithm (парсит таблицы в программы и исполняет их против утверждений). В 2021 году Aly et al. (2021) расширили датасет FEVER в новый набор

Claims in Social Media // In Working Notes of CLEF 2020. Conference and Labs of the Evaluation Forum (CLEF '20). CEUR-WS. – Thessaloniki. – Greece. – 2020.

³⁵ Vlachos A., Riedel S. Identification and Verification of Simple Claims about Statistical Properties // *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.* – 2015. – P. 2596–2601.

³⁶ Thorne J., Vlachos A. An Extensible Framework for Verification of Numerical Claims // *In European Chapter of the Association for Computational Linguistics (EACL).* – 2017. – P. 37–40.

³⁷ Gad-Elrab M. H., Stepanova D., Urbani J., Weikum G. Tracy: Tracing Facts over Knowledge Graphs and Text // *In Proceedings of the World Wide Web Conference (WWW '19). Association for Computing Machinery. – San Francisco. – USA. – 2019. – P. 3516–3520.*

³⁸ Shao C., Ciampaglia L. G., Flammini A., Menczer F. Hoaxy: A Platform for Tracking Online Misinformation // *In Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion). – International World Wide Web Conferences Steering Committee. – 2016. – P. 745–750.*

данных FEVEROUS, где некоторые из evidence были составлены на основе таблиц из Википедии. В отличие от TabFact для доказательств были дополнительно размечены нужные ячейки таблиц, а корпус из 16.000 таблиц был увеличен до 11.8 миллионов. В качестве бейзлайна для поиска информации по таблицам применялись классические методы поиска информации по тексту.

Определение позиции (stancedetection). Определение позиции утверждения по отношению к заданному тексту было целью задачи FakeNewsChallenge (Riedel et al., 2017; Hanselowski et al., 2018), а также задач RumourEval на SemEval в 2017 и 2019 годах (Derczynski et al., 2017; Gorrell et al., 2019). Позиция является важным фактором в задачах проверки фактов (объяснимый факт-чекинг) и обнаружении слухов наряду с лингвистической и мета-информацией. Ferreira and Vlachos (2016) предложили использовать позицию заголовка документа по отношению к утверждению во время проверки.

Демонстрационных систем и платформ при всем обилии задач относительно немного. Далее рассмотрены некоторые из них.

Noaxy (Shao et al., 2016) была представлена в 2016 году, как платформа для отслеживания дезинформации в социальных сетях и на новостных сайтах, включающей слухи, фейковые новости, искажения и т. п. Платформа собирает и отслеживает информацию из Twitter. Авторы считают, что она должна способствовать изучению динамики распространения фейковых и реальных новостей. Предварительные результаты их исследования показали, что дезинформация продвигается в основном за счет всего нескольких активных аккаунтов, после чего подхватывается СМИ.

CredEye (Popat et al., 2018) осуществляет оценку достоверности для анализа и обоснования наличия дезинформации в контенте. При оценке достоверности система ищет в сети релевантные документы при помощи поискового API Bing. Далее принимаются во внимание сразу несколько факторов: (1) соответствие языковых стилей, (2) позиция документа по отношению к утверждению и (3) степень достоверности источника. Позиция источника определяется моделью логистической регрессии над сниппетом, наиболее соответствующим утверждению по перекрытию биграмм. Степень доверия к источнику вычисляется с помощью подсчета количества случаев, в которых он поддерживал заслуживающую доверия и опровергал не заслуживающую доверия информацию.

Tracy (Gad-Elrab et al., 2019) является системой для проверки фактов на основе правил и графов знаний. В веб-интерфейсе системы пользователь может выбрать один из графов знаний и задать запрос вместе с набором правил. Система перефразирует данное на вход утверждение в набор более простых фактов с помощью rewriting модуля пока не может быть найдено подтверждение или опровержение для текущей формулировки. Затем на основе результатов проверки для полученного набора простых фактов логически строится вердикт для исходного утверждения. Таким образом в системе допускает использование нескольких источников для проверки.

AggChecker (Jo et al., 2019) является демо системой для проверки текстовых утверждений о численных агрегатах против базы данных. Каждое утверждение трансформируется в семантически эквивалентный SQL запрос. Авторы демонстрируют, что система позволяет добиться более высокой точности проверки текстовых утверждений, обобщающих информацию в таблицах, по сравнению со стандартными средствами SQL.

Scrutinizer (Karagiannis et al., 2020) была разработана для снижения затрат людей фактчекеров. Система переводит текстовые утверждения в SQL запросы над ассоциированной базой данных. Процесс контролируется оптимизатором, который учитывает взаимодействие с пользователем, то есть полезность отображенных в качестве доказательств или опровержений кандидатов и накладные расходы от их использования. Система помогает значительно снизить затраты на проверку фактов, при этом сохраняя исходное качество.

STANCY (Popat et al., 2019) проводит анализ утверждений, опираясь на разные точки зрения. Задача определения точки зрения рассматривается, как задача определения позиции (stance detection). Она решается при помощи сравнения векторных представлений утверждения и точки зрения, полученных при помощи сиамской вариации модели BERT. Одновременно модель пытается выучить позицию точки зрения и ограничения согласованности.

ФАКТА (Nadeem et al., 2019) осуществляет определение позиции извлеченных из Интернета документов по отношению к входному утверждению. Для поиска релевантных документов в Википедии и веб-источниках используется Google API, а для анализа отношения извлеченных документов к утверждению модель CNN над мешком слов (BoW). После этого система агрегирует результаты сравнения с каждым документом через усреднение предсказанных вероятностей классов «agree», «disagree» или «discuss» для получения итогового результата. Дополнительно система проводит лингвистический анализ, включающий анализ лексикона, предвзятых реплик, спорных слов, положительных/отрицательных настроений.

WhatTheWikiFact (Chernyavskiy et al., 2021) является демонстрацией системы для фактчекинга, основанной на постановке задачи FEVER. В отличие от ФАКТА она осуществляет проверку фактов с поиском доказательств на уровне предложений. В качестве результата сообщается вердикт и набор наиболее релевантных предложений с вероятностями позиций каждого из них к исходному утверждению. Система поддерживает несколько языков (русский, английский, болгарский). В пайплайне системы используется Google API над фразами группы существительного и именованными сущностями для поиска релевантных документов, затем TF-IDF для поиска релевантных предложений, далее языковая модель BERT для классификации каждого извлеченного предложения в паре с утверждением и последующая агрегация результатов моделью градиентного бустинга Catboost.

WikiCheck (Trokhymovych and Saez-Trumper, 2021) – API для фактчекинга утверждений против Википедии. Система так же как и WhatTheWikiFact базируется на датасете и формулировке задачи FEVER. Доказательства или опровержения ищутся через Wikipedia Search API, где в качестве запросов используются результаты NER парсера (именованные сущности). Затем для получения одной из итоговых меток SUPPORTS/REFUTES/NOT ENOUGH INFO применяется сиамская вариация модели BERT над парами вида <утверждение, доказательство>.

Tanbih (Zhang et al., 2019) является платформой, которая анализирует новостные статьи и средства массовой информации, а также прогнозирует фактологичность новостных сообщений, степень пропаганды, гиперпартийность, политическую предвзятость, фрейм и позицию к различным утверждениям и темам. Поддерживаются арабский и английский языки. Для анализа предвзятости и определения позиции используются подходы на основе BERT, а для определения фактуальности Support Vector Machine (SVM) над признаками, собранными с помощью сайта Media Bias/Fact Check (MBFC). Авторами Tanbih было представлено несколько демонстрационных систем:

Proppy для анализа пропаганды, то есть выявления предвзятых документов, целью которых является навязывание читателю конкретную позицию (Barrón-Cedeño et al., 2019). Для оценки степени пропаганды в системе используются различные признаки лексикона, стиля, загруженности текста, читаемости, предвзятости и векторные представления на основе TF-IDF.

ClaimRank для приоритизации фактов в документе, где в начале должны располагаться факты, которые нуждаются в проверке в первую очередь (Jaradat et al., 2018). Система использует двухслойную полносвязную нейронную сеть над вручную собранным набором признаков: TF-IDF, POS-теги, признаки лексикона, настроения и субъективности. Хотя модель была обучена только на политических текстах, по утверждению авторов она может быть применена к любым новостным документам.

FaxPlainAC (Zhang et al., 2021) – инструмент фактчекинга, который поддерживает циклическое взаимодействие между моделью и человеком. Авторы предполагают, что модель машинного обучения должна постоянно обновляться, учитывая отзывы и взаимодействие с пользователями. В отличие от рассмотренных ранее систем, данный инструмент собирает подробные сведения об истинности проверяемых утверждений и корректности предоставляемых доказательств. Для проверки сначала проводится отбор потенциально релевантных документов с помощью Google JSON API, затем модель на основе BERT осуществляет анализ входного утверждения в паре с каждым документом. Эта модель обучается одновременно предсказывать вердикт и извлекать из документа объяснения на основе данных FEVER.

Quin+ (Samarinas et al., 2021) – еще одна демо система объяснимого фактчекинга на основе датасета FEVER. Система сначала осуществляет поиск релевантных документов в вебе, затем поиск предложений и после этого проводит независимую классификацию утверждения в паре с извлеченными предложениями. Для получения итогового результата подсчитывается количество предсказанных меток Supports и Refutes. Особенностью системы является ее модуль для поиска релевантных доказательств. В его основе лежит система QR-BERT для кодирования фрагментов текста в векторные представления, и FAISS для поиска ближайших соседей в полученном векторном пространстве. Для поиска релевантных документов дополнительно применяется классический подход на основе мешка слов BM25, а для поиска предложений классификатор RoBERTa-large.

SciFact (Wadden et al., 2020) – система для проверки научных утверждений, извлекающая отрывки из исследовательской литературы в качестве доказательств или опровержений (корпуса S2ORC из публично доступных статей). В статье, посвященной данной системе, показано, что обучение и адаптация модели на предметную область может значительно повысить качество проверки по сравнению со стандартными моделями, обученными на корпусе статей из Википедии или политических текстах. Для демонстрационной системы авторы выбрали утверждения, связанные с COVID-19. Для поиска релевантных фрагментов текста в системе используется TF-IDF, а для поиска релевантных предложений и определения их отношения к утверждению модель BERT.

Заключение

Практически все представленные выше системы фактчекинга подразумевают либо отсутствие автоматизации и ручную проверку фактов специалистами с высоким уровнем знаний в области верификации информации, либо проприетарные и исследовательские системы верификации. В первом случае основными недостатками можно считать низкую скорость проверки информации и малый объём возможных проверок, а также обилие человеческих ошибок и влияние личной предвзятости проверяющего даже при использовании объективных методов верификации. Во втором случае практически все системы либо направлены на решение закрытых корпоративных задач и недоступны широкой публике, либо решают задачи исследований в области производства и распространения фейков, либо требуют специальных знаний (например, инструментов анализа данных, основ компьютерной лингвистики, языков программирования и т.д.) для получения результатов верификации. В целом простых в использовании систем, созданных для быстрой верификации информации онлайн рядовым пользователем, на английском языке единицы, а на русском языке не существует. Примером высокотехнологичных систем, созданных с учётом пользовательских сценариев для широкой аудитории, могут служить системы антиплагиата. На данный момент ни одна из онлайн-платформ фактчекинга не представляет сколько-нибудь завершённый проект с простым пользовательским интерфейсом, понятной широкой публике методологической основой верификации.

Кроме того, существующие системы фактчекинга опираются на верификацию информации с позиций аутентичности и на основании вторичных источников. В то же время практически не учитывается фактор пользовательского взаимодействия с информационными сообщениями, особенностей восприятия фейкового контента пользователями в разных платформах и средах, а также трансформацию информационных сообщений в контексте общения и пользовательского фидбэка.

Приложение. Обзор проектов и инициатив по применению и развитию фактчекинга

FactCheck.org (основан Annenberg Public Policy Center of the University of Pennsylvania, т.е. связан с ун-тами) – на сайте есть разделы с описанием разоблаченных фейков, по которым уже провели фактчекинг (в том числе разбирают посты из соцмедиа, в отдельный раздел выделены наиболее виральные истории), есть раздел с возможностью задать вопрос по фактчекингу или по вопросам государственной политики. Интересно также, что фейки про COVID-19 представлены и на испанском. Из дополнительных тем разбирают иммиграцию.

firstdraftnews.org (First Draft) – некоммерческая коалиция с девятью партнерами-учредителями, предоставляет практические и этические рекомендации по поиску, проверке и публикации контента из социальных сетей. В сентябре 2016 года стали международной партнерской сетью, состоящей из отделов новостей, университетов, платформ и организаций гражданского общества.

Poynter (www.poynter.org/ifcn) – некоммерческая школа журналистики и научно-исследовательская организация. Школа является владельцем газеты Tampa Bay Times и Международной сети проверки фактов, а также управляет PolitiFact.

PolitiFact (входит в Институт Пойнтера с 2018 г., имеет частных спонсоров) – это сайт, на котором установлен «измеритель правды», который определяет уровень правдивости политических заявлений (их шкала: правда, преимущественно правда, наполовину правда, преимущественно ложь, ложь, «врунишка!» (pants on fire)). Они имеют представительства в ряде штатов, в каждом из которых у них свои партнеры, проверяющие публичные высказывания властей местного уровня. В партнерах могут быть как НКО, так и СМИ (например, в штате Айова партнеры The Daily Iowan). На отдельной странице описывают методологию своей работы (<https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/>).

propagandacritic.com – ресурс по критике пропаганды, разбирает эффекты, при помощи которых реализуется пропаганда, предлагает гайд по распознаванию фейков (https://propagandacritic.com/index.php/tools-for-fighting-back/how-to-spot-fake-news/?utm_source=twitter&utm_medium=SAGE_social&utm_content=sagemedia_comm&utm_term=3d02fd10-02fe-4d14-8ce2-45b212d1ee81).

storyful.com - Компания предоставляет достоверные факты различным новостным организациям, таким как CNN, ABC News, Fox News, The New York Times. При проверке новости Storyful выясняет дату, место и личность автора материала, анализирует контент вплоть до мельчайших деталей: погоды, местности и даже звуков, если это теле- или радио-передача. Методологию не раскрывают (рекламная информация о том, как хорошо они работают), но есть блог и раздел с кейсами, где можно посмотреть примеры их работы (например, за 2019 год есть разбор с вышками 5G –<https://storyful.com/case-study/intelligence/5g-public-perception/>).

Расширения для браузеров и плагины

www.newsguardtech.com – сайт, а также расширение для браузера и мобильное приложение, которое оценивает качество новостных и информационных веб-сайтов и отслеживает дезинформацию в Интернете. Важно также, что у проекта есть коммерческое направление – есть возможность предоставления услуг для предприятий, включая инструмент безопасности бренда. Ранжирует сайты по:

- частоте публикации неточной информации;
- объему источников и очередности сообщения информации;
- степени разграничения между новостной и общественной журналистикой;

- точности заголовков, в том числе использование кликбейтов;
- степени раскрытия прав собственности на сайт, а также политические позиции владельцев.

trooclick.com – плагин для браузера предназначен для анализа текста и выявления неточностей. Trooclick сравнивает сообщение в тексте со статьями других СМИ. В случае несовпадения на экране высвечивается предупреждение, выявляя, таким образом, ресурсы, которым можно доверять с большей степенью вероятности.

Project Fib – расширение Chrome, которое утверждает, что обнаруживает фейковые новости в вашей ленте новостей Facebook и снимки в Твиттере. Если быть точнее, плагин маркирует верифицированные и не верифицированные посты, где вторые потенциально могут быть фейками.

Дополнительно можно отметить следующие проекты и инструменты:

tineye.com – инструмент помогает отследить, где и когда была сделана фотография, подвергалась ли она обработке, а также определит первоисточник картинки.

snopes.com. Цель ресурса – разоблачить или подтвердить городские легенды, слухи и сомнительные истории, скорее можно их обозначить как ресурс, который в большей степени работает с совсем неправдоподобными историями. В разделе методология не приводят ничего конкретного, поскольку «работают с разными кейсами – от определения, было ли отредактировано фото до проверки существования определенных документов», а потому каждый раз определяют ход работы поэтапно.

storyzy.com – международная группа из 8 человек, занимающаяся мониторингом фейков по миру на 17 языках. Дают демо-доступ к разделам «безопасность бренда» и «анализ данных и аналитика». Открыт раздел «медиаграмотность», где можно искать по отдельным статьям и видео на предмет того, насколько заслуживающими доверия являются источники, их публикующие, раздел с маркировкой сайтов, и раздел с образованием, но он преимущественно на французском.

Список использованной литературы

1. Alam F., Cresci S., Chakraborty T., Silvestri F., Dimitrov D., Da San Martino G., Shaar S., Firooz H., Nakov P. A Survey on Multimodal Disinformation Detection. – 2021. – URL: arXiv/2103.12541 (дата обращения 01.12.2021).
2. Allcott H. (2017) Social Media and Fake News in the 2016 Election // *Journal of Economic Perspectives*. – 2017. – № 31(2). – P. 211-236.
3. Alonso-Reina A., Sepulveda-Torres R., Saquete E., Palomar M. Team GPLSI. Approach for automated fact checking // In *Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER '19)*. Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 110–114.
4. Aly R., Guo Z., Schlichtkrull M., Thorne J., Vlachos A., Christodoulopoulos C., Cocarascu O., Mittal A. Feverous: Fact extraction and verification over unstructured and structured information. – 2021. – URL: arXiv:2106.05707 (дата обращения 01.12.2021).
5. Arnold M. An algorithm to detect fake news: A Q&A with Huan Liu and Kai Shu. – 2020. – URL: <https://research.asu.edu/algorithm-detect-fake-news-qa-huan-liu-and-kai-shu> (дата обращения 26.04.2021).
6. Baeza-Yates R. Bias on the Web // *Commun. ACM* 61. – 2018. – № 6. – P. 54–61.
7. Bakir V., McStay A. Fake news and the economy of emotions: problems, causes, solutions // *Digital Journalism*. – 2017. – № 6(2). – P. 1-22.
8. Barron-Cedeno A., Da San Martino G., Jaradat I., Nakov P. Proppy: A System to Unmask Propaganda in Online News. – 2019. – URL: ArXiv abs/1912.06810 (дата обращения 01.12.2021).
9. Barron-Cedeno A., Elsayed T., Nakov P., Da San Martino G., Hasanain M., Suwaileh R., Haouari F., Babulkov N., Hamdan B., Nikolov A., Shaar S., Ali Z. S. Overview of CheckThat! – Automatic Identification and Verification of Claims in Social Media // In *Proceedings of the 11th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF '20)*. – Thessaloniki. – Greece. – 2020. – P. 215–236.
10. Barron-Cedeno A., Elsayed T., Nakov P., Da San Martino G., Hasanain M., Suwaileh R., Haouari F. CheckThat! at CLEF 2020: Enabling the Automatic Identification and Verification of Claims in Social Media // In *Proceedings of the 42nd European Conference on Information Retrieval (ECIR '20)*. – Lisbon. – Portugal. – 2020. – P. 499–507.
11. Bermudez J. P. The Post-Truth temperament: What makes belief stray from evidence? And what can bring them back together? // *America's Post-Truth Phenomenon: When Feelings and Opinions Trump Facts and Evidence* / C.G. Prado (ed.). – 2018. – P. 87-109.
12. Calzada, J., Gil, R. What Do News Aggregators Do? Evidence from Google News in Spain and Germany // *Marketing Science*. 2020. – № 39(1). – P. 134-167.
13. Castillo C., Mendoza M., Poblete B. Information Credibility on Twitter // In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)* / Association for Computing Machinery. – Hyderabad. – India. – 2011. – P. 675–684.
14. Chakrabarty T., Alhindi T., Muresan S. Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification // In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18)*. Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 127– 131.
15. Chen W., Wang H., Chen J., Zhang Y., Wang H., Li S., Zhou X., Yang Wang. W. TabFact: A Large-scale Dataset for Table-based Fact Verification // In *Proceedings of the 8th International Conference on Learning Representations (ICLR '20)*. – Addis Ababa. – Ethiopia. 2020.
16. Chernyavskiy A., Ilvovsky D., Nakov P. What The Wiki Fact: Fact-Checking Claims Against Wikipedia // *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery. – New York. – USA. – 2021. – P. 4690–4695.

17. Ciampaglia G. L., Shiralkar P., Rocha L. M., Bollen J., Menczer F., Flammini A. Computational fact checking from knowledge networks // PLoS ONE. – 2015. – № 10(6). – P. 1–13.
18. Comai G. Responding to alleged Russian interference by focusing on the vulnerabilities that make it possible // Democracy and Fake News. Information Manipulation and Post-Truth Politics / ed. Giusti S., Piras E. – London: Routledge. – 2021. – P. 143-154.
19. Conroy N. J., Rubin V. L., Chen Y. Automatic deception detection: Methods for finding fake news // Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. – 2016. – P. 1-4.
20. Da San Martino G., Cresci S., Barron-Cedeno A., Yu S., Di Pietro R., Nakov P. A Survey on Computational Propaganda Detection // In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-PRICAI '20). International Joint Conferences on Artificial Intelligence Organization. – Yokohama. – Japan. – 2020. – P. 4826–4832.
21. Dauce F., Loveluck B. Codes of conduct for algorithmic news recommendation: The Yandex.News controversy in Russia. – 2021. – First Monday. – № 26(5).
22. Derakhshan H., Wardle C. Information Disorder: Definitions. Understanding and Addressing the Disinformation Ecosystem. Pennsylvania: Annenberg School for Communication. – 2017. – P. 5-12. – URL: <https://firstdraftnews.org/wp-content/uploads/2018/03/The-Disinformation-Ecosystem-20180207-v2.pdf> (дата обращения 26.04.2021).
23. Derczynski L., Bontcheva K., Liakata M., Procter R., Wong Sak Hoi G., Zubiaga A. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours // In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval '17). Association for Computational Linguistics. – Vancouver. – Canada. – 2017. – P. 69–76.
24. Eurobarometer. Final results of the Eurobarometer on fake news and online disinformation. – 2018. – URL: <https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation> (дата обращения 26.04.2021).
25. European Commission Guidance on Strengthening the Code of Practice on Disinformation // Brussels: EC. – 2021. – URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/strengthening-eu-code-practice-disinformation_en (дата обращения 01.12.2021).
26. Fairfield P. Lords of mendacity // America's Post-Truth Phenomenon: When Feelings and Opinions Trump Facts and Evidence / C.G. Prado (ed.). – 2018. – CA: Praeger. – P. 153-164.
27. Fake News Net // Fake News Tracker. – 2019. – URL: <https://github.com/KaiDMML/FakeNewsNet> (дата обращения 26.04.2021)
28. Farkas J., Schou J. Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. Javnost // The Public. – 2018. – № 25(3). – P. 298-314.
29. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. The Rise of Social Bots // Commun. ACM 59. – 2016. – № 7. – P. 96–104.
30. Ferreira W., Vlachos A. Emergent: a novel data-set for stance classification // In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – San Diego. – California. – 2016. – P. 1163–1168.
31. Fu H., Li H., Liu Y., Pirkkalainen H., Salo M. Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload // Information Processing & Management. – 2020. – №57 (6).
32. Gad-Elrab M. H., Stepanova D., Urbani J., Weikum G. Tracy: Tracing Facts over Knowledge Graphs and Text // In Proceedings of the World Wide Web Conference (WWW '19). Association for Computing Machinery. – San Francisco. – USA. – 2019. – P. 3516–3520.
33. Gelfert A. Fake News: A Definition // Informal Logic. – 2018. – № 38 (1). – P. 84–117.
34. Gorrell G., Kochkina E., Liakata M., Aker A., Zubiaga A., Bontcheva K., Derczynski L. SemEval-2019 Task 7: Rumour Eval, Determining Rumour Veracity and Support for Rumours //

- In Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval '19). Association for Computational Linguistics. – Minneapolis. –USA. – 2019. – P. 845–854.
35. Guadagno R.E., Guttieri K. Fake News and Information Warfare: An Examination of the Political and Psychological Processes From the Digital Sphere to the Real World // Handbook of Research on Deception, Fake News, and Misinformation Online / E. Chilwa, S.A. Samoilenko (eds.). – Hershey PA: IGI Global. – 2019. – P. 167–191.
 36. Hanselowski A., Avinesh PVS, Schiller B., Caspelherr F., Chaudhuri D., Meyer C. M., Gurevych I. A Retrospective Analysis of the Fake News Challenge Stance-Detection Task // In Proceedings of the 27th International Conference on Computational Linguistics (COLING '18). Association for Computational Linguistics. – Santa Fe. – New Mexico. – USA. – 2018. –P. 1859–1874.
 37. Hanselowski A., Zhang H., Li Z., Sorokin D., Schiller B., Schulz C., Gurevych I. UKP-Athene: Multi-Sentence Textual Entailment for Claim Verification // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18). Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 103–108.
 38. Hardalov M., Arora A., Nakov P., Augenstein I. A Survey on Stance Detection for Mis- and Disinformation Identification. – 2021. – URL: [arXiv/2103.00242](https://arxiv.org/abs/2103.00242) (дата обращения 01.12.2021).
 39. Harsin J. A Critical Guide to Fake News: from Comedy to Tragedy (translated from «Un guide critique des fake news: de la comédie à la tragédie») // Pouvoirs: Revue d'Etudes Constitutionnelles et Politiques. – 2018. – № 164(1). – P. 99–119.
 40. Hasanain M., Haouari F., Suwaileh R., Sheikh Ali Z., Hamdan B., Elsayed T., Barrón-Cedeño A., Da San Martino G., Nakov P. Overview of CheckThat! 2020 Arabic: Automatic Identification and Verification of Claims in Social Media // In Working Notes of CLEF 2020—Conference and Labs of the Evaluation Forum (CLEF '2020). CEUR-WS.org. – Thessaloniki. – Greece. – 2020.
 41. Hidey C., Diab M. Team SWEEPer: Joint Sentence Extraction and Fact Checking with Pointer Networks // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18). – Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 150–155.
 42. Jaradat I., Gencheva P., Barrón-Cedeño A., Márquez L., Nakov P. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (NAACL '18). Association for Computational Linguistics. – New Orleans. – Louisiana. – 2018. – P. 26–30.
 43. Jo S., Trummer I., Yu W., Wang X., Yu C., Liu D., Mehta N. Agg Checker: a fact-checking system for text summaries of relational data sets // Proc. VLDB Endow. – 2019. – № 12. – P. 1938–1941.
 44. Karagiannis G., Saeed M., Papotti P., Trummer I. Scrutinizer: A Mixed-Initiative Approach to Large-Scale, Data-Driven Claim Verification // Proc. VLDB Endow. – 2020. – № 13, 12. – P. 2508–2521.
 45. Khaldarova I., Pantti M. Fake News // Journalism Practice. – 2016. – № 10(7). – P. 891–901.
 46. Krasni J. How to hijack a discourse? Reflections on the concepts of post-truth and fake news // Humanities and social sciences communications. – 2020. – № 7 (23). – P. 1-9. – URL: <https://doi.org/10.1057/s41599-020-0527-z> (дата доступа 26.04.2021).
 47. Küçük D., Can F. Stance Detection // ACM Comput. Surv. – 2020. – № 53, 1. – P. 37.
 48. Lazer D. M. J., Baum M. A., Benkler Y., Berinsky A. J., Greenhill K. M., Menczer F., Metzger M. J., Nyhan B., Pennycook G., Rothschild D., Schudson M., Sloman S. A., Sunstein C. R., Thorson E. A., Watts D.J., Zittrain J. L. The science of fake news // Science. – 2018/ – 359, 6380. – P. 1094–1096.
 49. Lee N., Bang Y., Madotto A., Fung P. Towards Few-shot Fact-Checking via Perplexity // In Proceedings of the 2021 Conference of the North American Chapter of the Association for

- Computational Linguistics: Human Language Technologies (NAACL-HLT '21). Association for Computational Linguistics. – Online. – 2021. – P. 1971–1981.
50. Lee N., Li B., Wang S., Yih Wen-tau, Ma H., Khabsa M. Language Models as Fact Checkers? // In Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER '20). Association for Computational Linguistics. – Online. – 2020. – P. 36–41.
51. Li Y., Gao J., Meng C., Li Q., Su L., Zhao B., Fan W., Han J. A Survey on Truth Discovery // SIGKDD Explor. Newsl. – 2016. – № 17, 2. – P. 1–16.
52. Malon C. Team Papelo: Transformer Networks at FEVER // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18). – Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 109–113.
53. McIntyre L. Post-truth. – Cambridge, MA: MIT Press. – 2018.
54. Meese J., Hurcombe E. Facebook, News Media and Platform Dependency: The Institutional Impacts of News Distribution on Social Platforms // New Media & Society. – 2020. – URL: <https://journals.sagepub.com/doi/full/10.1177/1461444820926472> (дата обращения 01.12.2021).
55. Nadeem M., Fang W., Xu B., Mohtarami M., Glass J. FAKTA: An Automatic End-to-End Fact Checking System // In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations) (NAACL-HLT '19). Association for Computational Linguistics. – Minneapolis – Minnesota. – 2019. – P. 78–83.
56. Nakov P., Corney ., Hasanain M., Alam F., Elsayed T., Barrón-Cedeño A., Papotti P., Shaar S., Da San Martino G. Automated Fact-Checking for Assisting Human Fact-Checkers // In Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI '21). International Joint Conferences on Artificial Intelligence Organization. – Online. – 2021. – P. 4551–4558.
57. Nakov P., Da San Martino G., Elsayed T., Barrón-Cedeño A., Míguez R., Shaar S., Alam F., Haouari F., Hasanain M., Babulkov N., Nikolov A., Shahi G. K., Struß J. M., Mandl T. The CLEF-2021 CheckThat! Lab on Detecting CheckWorthy Claims, Previously Fact-Checked Claims, and Fake News // In Proceedings of the 43rd European Conference on Information Retrieval (ECIR '21). – Italy. – 2021. – P. 639–649.
58. Nakov P., Da San Martino G., Elsayed T., Barrón-Cedeño A., Míguez R., Shaar S., Alam F., Haouari F., Hasanain M., Mansour W., Hamdan B., Sheikh Ali Z., Babulkov N., Nikolov Alex, Kishore Shahi G., Struß J. M., Mandl T., Kutlu M., Kartal Y. S. Overview of the CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News // In Proceedings of the 12th International Conference of the CLEF Association: Information Access Evaluation Meets Multilinguality, Multimodality, and Visualization (CLEF '21). – Bucharest. – Romania. – 2021.
59. Nakov P., Sencar H. T., An J., Kwak H. A Survey on Predicting the Factuality and the Bias of News Media. – 2021. – URL: arXiv/2103.12506 (дата обращения 01.12.2021).
60. News in Spain and Germany. – 2017. – June 20. – URL: https://www.nes.ru/dataupload/files/science/Research%20seminars/google_news_june_2017.pdf (дата обращения 26.04.2021).
61. Nie Y., Chen H., Bansal M. Combining Fact Extraction and Verification with Neural Semantic Matching Networks // In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '19). – 2019. – P. 6859–6866.
62. Otto W. 2. Team GESIS Cologne: An all in all sentence-based approach for FEVER // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18) / Association for Computational Linguistics. – 2018. – Brussels. – Belgium. – P. 145–149.
63. Pariser E. The Filter Bubble: What the Internet Is Hiding from You. – New York: Penguin Press. – 2011.
64. Petroni F., Rocktäschel T., Riedel S., Lewis P., Bakhtin A., Wu Y., Miller A. Language Models as Knowledge Bases? // In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language

- Processing (EMNLP-IJCNLP '19). Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 2463–2473.
65. Popat K., Mukherjee S., Strötgen J., Weikum G. CredEye: A Credibility Lens for Analyzing and Explaining Misinformation // In Companion Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee. – Lyon. – France. – 2018. – P. 155–158.
 66. Popat K., Mukherjee S., Yates A., Weikum G. STANCY: Stance Classification Based on Consistency Cues // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP/IJCNLP '19). Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 6413– 6418.
 67. Posetti J., Matthews A. A short guide to the history of 'fake news' and disinformation. A Learning Module for Journalists and Journalism Educators. – 2018. – Washington: International Center for Journalists. – URL: https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf (дата обращения 26.04.2021).
 68. Rashkin H., Choi E., Jang Jin Yea, Volkova S., Choi Y. Truth of Varying Shades: Analyzing Language in Fake News and Political FactChecking // In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP '17). Association for Computational Linguistics. – Copenhagen. – Denmark. – 2017. – P. 2931–2937.
 69. Reddy A. J., Rocha G., Esteves D. DeFactoNLP: Fact Verification using Entity Recognition, TFIDF Vector Comparison and Decomposable Attention // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18). Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 132–137.
 70. Riedel B., Augenstein I., Spithourakis G. P., Riedel S. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. – 2017. – URL: ArXiv abs/1707.03264 (дата обращения 26.04.2021).
 71. Rini R. Fake news and partisan epistemology // Kennedy Institute of Ethics Journal. – 2017. – № 27(2). – P. 43-64.
 72. Roberts A., Raffel C., Shazeer N. How Much Knowledge Can You Pack Into the Parameters of a Language Model? // In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP '20). Association for Computational Linguistics. – 2020. – Online. – P. 5418–5426.
 73. Samarin C., Hsu W., Lee M. L. Proceedings of the 2021 // Conference of the North American. Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations. Association for Computational Linguistics. – 2021. – P. 84–91.
 74. Shaar S., Babulkov N., Da San Martino G., Nakov P. That is a Known Lie: Detecting Previously Fact-Checked Claims // In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL '20). Association for Computational Linguistics. – Online. – 2020. – P. 3607–3618.
 75. Shaar S., Haouari F., Mansour W., Hasanain M., Babulkov N., Alam F., Da San Martino G., Elsayed T., Nakov P. Overview of the CLEF-2021 CheckThat! Lab Task 2 on Detecting Previously Fact-Checked Claims in Tweets and Political Debates // In Working Notes of CLEF 2021. Conference and Labs of the Evaluation Forum (CLEF '21). CEUR-WS. – Bucharest. – Romania. – 2021.
 76. Shaar S., Nikolov A., Babulkov N., Alam F., BarronCedeno A., Elsayed T., Hasanain M., Suwaileh R., Haouari F., Da San Martino G., Nakov P. Overview of CheckThat! 2020 English: Automatic Identification and Verification of Claims in Social Media // In Working Notes of CLEF 2020. Conference and Labs of the Evaluation Forum (CLEF '20). CEUR-WS. – Thessaloniki. – Greece. – 2020.
 77. Shao C., Ciampaglia L. G., Flammini A., Menczer F. Hoaxy: A Platform for Tracking Online Misinformation // In Proceedings of the 25th International Conference Companion on

- World Wide Web (WWW '16 Companion). – International World Wide Web Conferences Steering Committee. – 2016. – P. 745–750.
78. Shu K., Mahudeswaran D., Wang S., Lee D., Liu H. Fake News Net: A Data Repository with News Content, Social Context and Spatial temporal Information for Studying Fake News on Social Media. – Cornell University. – 2018. – URL: <https://arxiv.org/abs/1809.01286> (дата обращения 26.04.2021).
 79. Shu K., Sliva A., Wang S., Tang J., Liu H. Fake News Detection on Social Media: A Data Mining Perspective // SIGKDD Explor. Newsl. – 2017. – № 19, 1. – P. 22–36.
 80. Silverman C., Singer-Vine J., Vo L. Th. In Spite Of The Crackdown, Fake News Publishers Are Still Earning Money From Major Ad Networks. – 2017. – URL: <https://www.buzzfeednews.com/article/craigsilverman/fake-news-real-ads> (дата обращения 26.04.2021).
 81. Silverman C., Singer-Vine J., Vo L. Th. In Spite Of The Crackdown, Fake News Publishers Are Still Earning Money From Major Ad Networks // Buzz Feed News. – 2017. – URL: <https://www.buzzfeednews.com/article/craigsilverman/fake-news-real-ads> (дата обращения 26.04.2021).
 82. Somers M. Deepfakes, explained // MIT Sloan. – 2020. – URL: https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained?utm_source=mithp&utm_medium=web&utm_campaign=deepfakeexplainer (дата обращения 26.04.2021).
 83. Stambach D., Neumann G. Team DOMLIN: Exploiting Evidence Enhancement for the FEVER Shared Task // In Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER '19). Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 105–109.
 84. Steinberg L. Infographic: Beyond Fake News – 10 Types of Misleading News. – 2017. – URL: <https://eavi.eu/beyond-fake-news-10-types-misleading-info/> (дата обращения 26.04.2021).
 85. Tandoc E. C., Lim Z. W., Ling R. Defining Fake News: A typology of scholarly definitions // Digital Journalism. – 2017. – № 6(3). – P. 1–17.
 86. Thorne J., Vlachos A. An Extensible Framework for Verification of Numerical Claims // In European Chapter of the Association for Computational Linguistics (EACL). – 2017. – P. 37–40.
 87. Thorne J., Vlachos A. Automated Fact Checking: Task Formulations, Methods and Future Directions // In Proceedings of the 27th International Conference on Computational Linguistics (COLING '18). Association for Computational Linguistics. – Santa Fe. – New Mexico. – USA. – 2018. – P. 3346–3359.
 88. Thorne J., Vlachos A., Christodoulopoulos C., Mittal A. FEVER: a Large-scale Dataset for Fact Extraction and VERification // In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT '18). Association for Computational Linguistics. – New Orleans. – Louisiana. – USA. – 2018. – P. 809–819.
 89. Thorne J., Vlachos A., Cocarascu O., Christodoulopoulos C., Mittal A. The Fact Extraction and VERification (FEVER) Shared Task // In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER '18). Association for Computational Linguistics. – Brussels. – Belgium. – 2018. – P. 1–9.
 90. Thorne J., Vlachos A., Cocarascu O., Christodoulopoulos C., Mittal A. The FEVER2.0 Shared Task // In Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER '19). Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 1–6.
 91. Trokhymovych M., Saez-Trumper D. WikiCheck: An End-to-end Open Source Automatic Fact-Checking API based on Wikipedia // In Proceedings of the 30th ACM International

- Conference on Information & Knowledge Management (CIKM '21). Association for Computing Machinery. – New York. – USA. – 2021. – P. 4155–4164.
92. Ulken E. Non-traditional sources cloud Google News results // Online Journalism. – URL: <http://www.ojr.org/050519ulken> (дата обращения 26.04.2021).
 93. Vamanu I. Fake News and Propaganda: A Critical Discourse Research Perspective // Open Information Science. 2019. – № 3(1). – P. 197–208.
 94. Van Dijck J., Nieborg D., Poell, T. Reframing platform power // Internet Policy Review. – 2019. – № 8(2).
 95. Vlachos A., Riedel S. Identification and Verification of Simple Claims about Statistical Properties // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. – 2015. – P. 2596–2601.
 96. Vosoughi S., Roy D., Aral S. The spread of true and false news online // Science – 2018. – P. 1146–1151.
 97. Wadden D., Lin S., Lo K., Wang L. L., Zuylen M., Cohan A., Hajishirzi H. Fact or Fiction: Verifying Scientific Claims // In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP '20). – Online. – 2020. – P. 7534–7550.
 98. Wendling M. The (almost) complete history of “fake news” // BBC Trending. – 2018. – 22 January. – URL: www.bbc.com/news/blogs-trending-42724320 (дата доступа 01.12.2021).
 99. Wiggins B.E. Navigating an immersive narratology: Factors to explain the reception of fake news // International Journal of E-Politics (IJEP). – 2017. – № 8 (3). – P. 16-29.
 100. Wijermars, M. Russia’s law «On news aggregators»: Control the news feed, control the news? // Journalism. – 2021. – URL: <https://doi.org/10.1177/1464884921990917> (дата доступа 01.12.2021).
 101. Yang Wang W. «Liar, Liar Pants on Fire»: A New Benchmark Dataset for Fake News Detection // In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL '17). Association for Computational Linguistics. – Vancouver. – Canada. – 2017. – P. 422–426.
 102. Yifan Z, Da San Martino G., Barron-Cedeno A., Romeo S., An J., Kwak H., Staykovski T., Jaradat I., Karadzhov G., Baly R., Darwish K., Glass J., Nakov P. Tanbih: Get To Know What You Are Reading // In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP '19): System Demonstrations. Association for Computational Linguistics. – Hong Kong. – China. – 2019. – P. 223–228.
 103. Zannettou S., Sirivianos M., Blackburn J., Kourtellis N. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans J. // Data and Information Quality. – 2019. – № 11, 3. – P. 10:1–10:3.
 104. Zaryan S. Truth and Trust: How audiences are making sense of Fake News. – MSc Media and Communication: Lund University – 2017.
 105. Zhang Z., Rudra K., Anand A. FaxPlainAC: A Fact-Checking Tool Based on EXPLAINable Models with HumAn Correction in the Loop // Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Association for Computing Machinery. – New York. – USA. – 2021. – P. 4823–4827.
 106. Ziyu Y. Weibo Gives Media, Government Power to Quash «Rumors» // Sixth Tone. – 2018. – URL: <http://www.sixthtone.com/news/1003152/weibo-gives-media,-government-power-to-quash-rumors> (дата обращения 26.04.2021).
 107. Zubiaga A., Aker A., Bontcheva K., Liakata M. Procter Rob Detection and Resolution of Rumours in Social Media // ACM Comput. Surv. – 2018. – № 51, 2. – P. 36.
 108. Айртон Ш., Позетти Дж. Журналистика, «фейковые новости» и дезинформация. – URL: <https://unesdoc.unesco.org/ark:/48223/pf0000371785/PDF/371785rus.pdf.multi> (дата обращения 26.04.2021).

109. Архипова А. Александра Архипова об инфодемии, фейк-ньюс и практиках солидаризации. – 2020. – URL: <http://researcharts.ru/arkhipova-infodemic> (дата обращения 26.04.2021).
110. Гатов В. FakeNews: правда о неправде. – URL: <https://www.ridl.io/ru/fake-news-pravda-o-nepravde> (дата обращения 26.04.2021).
111. Гид Университета Огайо: Ohio University MPA Fake News, Misinformation, & Fact-Checking. – URL: <https://onlinemasters.ohio.edu/masters-public-administration/guide-to-misinformation-and-fact-checking/> (дата обращения 26.04.2021).
112. Гордеев В. Бастрыкин предложил «вакцинировать» право для борьбы с фейками // РБК. – 2021. – 17 мая. – URL: <https://www.rbc.ru/politics/17/05/2021/60a271af9a79476afde3b2b6> (дата обращения 01.12.2021).
113. Левченко Г. Главы Twitter, Google и Facebook приняли участие в онлайн-слушаниях в Сенате США, посвященных модерированию контента. Цукерберг не сразу дозвонился // Медуза. – 2020. – URL: <https://meduza.io/feature/2020/10/28/glavy-twitter-google-i-facebook-prinyali-uchastie-v-onlayn-slushaniyah-v-senate-ssha-posvyaschennyh-moderirovaniyu-kontenta-tsukerberg-ne-srazu-dozvonilsya> (дата обращения 26.04.2021).
114. Нормы сообщества. Часть IV. Целостность и подлинность // Facebook. – 2020. URL: https://www.facebook.com/communitystandards/integrity_authenticity (дата обращения 26.04.2021).
115. Петров Н., Архипова А., Спиридонов В., Пейгин Б. Инфодемия: существующие подходы к анализу паник, фобий, слухов, фейков во время эпидемий и предложения по борьбе с ними // Мониторинг экономической ситуации в России. Тенденции и вызовы социально-экономического развития. – 2020. – № 8(110). – С. 80–87.
116. Политика в отношении манипуляции платформой и спама // Twitter. – 2020. – URL: <https://help.twitter.com/ru/rules-and-policies/platform-manipulation> (дата обращения 26.04.2021).
117. Почепцов Г. Фейки и дезинформация: в чем их сила? // Хвиля. – 2018. – URL: <https://hvylya.net/analytics/society/feyki-i-dezinformatsiya-v-chem-ih-sila.html> (дата обращения 26.04.2021).
118. Правила и безопасность // Youtube. – 2020. – URL: <https://www.youtube.com/intl/ru/about/policies/#community-guidelines> (дата обращения 26.04.2021).
119. Скрынникова А., Скобелев В. Роспечать предложила создать сервис для распознавания фейковых новостей // РБК. – 2020. – URL: Доступно по ссылке: https://www.rbc.ru/technology_and_media/04/08/2020/5f2832869a7947d218ad4dd9 (дата обращения 26.04.2021).
120. Хвостик Е. Интернет-компании хотят больше свободы в удалении нелегального контента // Коммерсант. – 2020. – URL: <https://www.kommersant.ru/doc/4548672> (дата обращения 26.10.2020).
121. Шерилин А., Позетти Д. Журналистика, «фейковые новости» и дезинформация // ЮНЕСКО. – 2018. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000371785/PDF/371785rus.pdf.multi> (дата обращения 26.04.2021).
122. Яндекс.Дзен. – URL: <https://yandex.ru/support/zen/requirements/fact-checking.html> (дата обращения 26.04.2021).