

Вестник Череповецкого государственного университета. 2022. № 6 (111). С. 89–109.
Cherepovets State University Bulletin, 2022, no. 6 (1101), pp. 89–109.

Научная статья

УДК 658.382.3:316.64

<https://doi.org/10.23859/1994-0637-2022-6-111-7>

Моделирование процессов распространения дезинформации в информационных сетях

Анатолий Николаевич Швецов^{1✉}, Сергей Владимирович Дианов²,
Анастасия Александровна Попова³, Дмитрий Олегович Кузнецов⁴

^{1, 2, 3, 4} Вологодский государственный университет,
Вологда, Россия,

¹smithv@mail.ru✉, <https://orcid.org/0000-0002-7021-5184>

²dianov.sv@mail.ru, <https://orcid.org/0000-0001-8297-8077>

³afitskevich@bk.ru

⁴awesomeshout@yandex.ru

Аннотация. Рассматривается проблема распространения дезинформации в информационно-телекоммуникационных сетях. Анализируются методы обнаружения дезинформации, обсуждаются информационные и социальные эффекты, способствующие распространению недостоверной информации в сети Интернет. Предлагается субъектно-реляционная модель процесса передачи дезинформации, представленная полностью связанным ориентированным графом, связывающим основные классы возможных субъектов. Вводятся понятия графа и матрицы информационных связей агентов, определяется модель агента процесса распространения дезинформации на основе вероятностного автомата, описывающего поведение агента в среде Интернета.

Ключевые слова: информационно-телекоммуникационные системы, интернет-среда, агент процесса распространения дезинформации

Благодарность. Статья выполнена из средств субсидии федерального бюджета на разработку программ бакалавриата и программ магистратуры по профилю «искусственный интеллект», а также на повышение квалификации педагогических работников образовательных организаций высшего образования в сфере искусственного интеллекта (соглашение №075-15-2021-1039 от 28.09.2021).

Для цитирования: Швецов А. Н., Дианов С. В., Попова А. А., Кузнецов Д. О. Моделирование процессов распространения дезинформации в информационных сетях // Вестник Череповецкого государственного университета. 2022. № 6 (111). С. 89–109. <https://doi.org/10.23859/1994-0637-2022-6-111-7>.

Modeling the spread of disinformation in information networks

Anatolii N. Shvetsov^{1✉}, Sergey V. Dianov²,
Anastasia A. Popova³, Dmitry O. Kuznetsov⁴

^{1,2,3,4} Vologda State University,
Vologda, Russia,

¹smithv@mail.ru✉, <https://orcid.org/0000-0002-7021-5184>

²dianov.sv@mail.ru, <https://orcid.org/0000-0001-8297-8077>

³afitskevich@bk.ru

⁴awesomeshout@yandex.ru

Abstract. The article considers the problem of disinformation spread in information and telecommunication networks. The authors analyze the methods for detecting disinformation; discuss informational and social effects that contribute to the dissemination of false information on the Internet. A subject-relational model of the disinformation transfer process is proposed, represented by a fully connected directed graph that connects the main classes of possible subjects. The article introduces concepts of a graph and a matrix of information links of agents, defines an agent model of the disinformation spread process based on a probabilistic automaton that describes the behavior of an agent in the Internet environment.

Keywords: information and telecommunication systems, Internet environment, agent of disinformation spread process

Acknowledgments. The article was funded by the federal budget subsidy for the development of bachelor's and master's programs of "Artificial intelligence"; for the advanced training of lecturers of higher education organizations in the field of artificial intelligence (agreement no. 075-15-2021-1039 dated 28.09.2021).

For citation: Shvetsov A. N., Dianov S. V., Popova A. A., Kuznetsov D. O. Modeling the spread of disinformation in information networks. *Cherepovets State University Bulletin*, 2022, no. 6 (111), pp. 89–109 (In Russ.). <https://doi.org/10.23859/1994-0637-2022-6-111-7>.

Введение

Виртуально-информационная реальность, формируемая современными информационно-телекоммуникационными сетями, приобретает формы быстроразвивающихся цифровых миров, захватывающих расширяющиеся пространства материальной действительности, организационной и производственной среды, наполняет значительную часть сознания человека, поглощает растущие объемы его физического времени, формирует новые психологические ценности и социальные установки, оказывая растущее влияние на характер производственных процессов, формы трудовых отношений, социальное поведение членов общества.

Информационно-коммуникационные процессы в современном обществе порождают проблемы анализа характера и последствий информационных воздействий на массовое сознание и психологическое состояние отдельного человека посредством глобальных информационно-телекоммуникационных систем и сервисов (Интернет, социальные сети, сетевые сообщества, телеграмм-каналы и т. п.). Такие информационные воздействия открывают широкие возможности манипуляции общественным сознанием, активного продвижения дезинформации, формирования деструктивного поведения некоторых категорий населения.

В условиях глобального информационного общества феномен дезинформации

становится фундаментальным фактором экономических, политических и социальных отношений, проявляется в самых различных формах жизнедеятельности индивидов, коллективных и общественных структур, институтов муниципального, регионального и государственного управления.

Согласно данным отчета Digital 2022 July Global Statshot Report¹, число пользователей Интернета в мире достигло 5,03 млрд чел., что составило 63,1 % населения Земли, количество активных пользователей социальных сетей оценивается в 4,70 млрд чел., т. е. более половины жителей планеты. В среднем каждый пользователь социальных сетей проводит в них 2 часа 29 минут ежедневно.

В соответствии с концепцией Д. И. Дубровского² авторы рассматривают обмен как сложный многостадийный информационно-психологический процесс передачи, обработки и оценки информационных сообщений, принятия решений, вытекающих из полученных при обработке и оценке результатов, осуществления действий, обусловленных принятыми решениями, субъектами данного процесса. Такими субъектами могут быть отдельные люди; коллективы людей, связанные между собой различными отношениями – родственными, деловыми, идеологическими; большие социальные группы, классы, народы, образующие массового субъекта; социальные институты общества, образующие институциональные субъекты.

Также в данном исследовании авторы опираются на современные определения и трактовки феномена дезинформации, сформулированные в работах, определяющих это понятие как объект³ и результат⁴ тщательно спланированного и технически эффективно организованного процесса обмана⁵. В данной публикации авторы рассматривают дезинформацию как осозанный, целенаправленный обман, вводящий получателя информации в заблуждение (disinformation), отличая ее от неполной, неточной или недостоверной информации, возникающей непреднамеренно, за которой не стоит злой умысел отправителя информации (misinformation). Дезинформация способна оказывать значительное влияние на государственную политику⁶, общественное мнение⁷ и приводить к серьезным социальным⁸ и экономическим последствиям¹.

¹ Digital 2022 July Global Statshot Report. URL: <https://datareportal.com/reports/digital-2022-july-global-statshot?rq=Digital%202022/> (дата обращения 01.09.2022).

² Дубровский Д. И. Обман. Философско-психологический анализ. Москва: Конон+, 2010. 336 с.

³ Fallis D. The Epistemic Threat of Deepfakes // *Philosophy&Technology*. 2021. Vol. 34. P. 623–643.

⁴ Soe S. O. A Unified Account of Information, Misinformation, and mation / In: *Synthese – An international journal for epistemology, methodology and philosophy of science*. 2021. Vol. 198. P. 5929–5949.

⁵ Kumar S., West R., Lescovec J. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes // *WWW '16: Proceedings of the 25th International Conference on World Wide Web April 2016*. P. 591–602. URL: <https://doi.org/10.1145/2872427.2883085>.

⁶ Колин К. К. Философские проблемы информатики. Москва: БИНОМ, 2010. 264 с.

⁷ История информатики и философия информационной реальности / под редакцией Р. М. Юсупова, В. П. Котенко. Москва: Академический проект, 2007. 430 с.

⁸ Colic-Peisker V., Flitney A. The Age of Post-Rationality: Limits of economic reasoning in the 21st century. Singapore: Palgrave MacMillan, 2018. 253 p.

Социальные сети принципиально изменили возможности взаимодействия людей в глобальном информационном пространстве. В начале 90-х годов прошлого века английский антрополог и эволюционный психолог Робин Данбар (Robin Ian Mac-Donald Dunbar) ввел в научный оборот понятие «число Данбара»², определяющее количество постоянных устойчивых социальных связей, которые способен поддерживать индивид (в нашем случае Homo Sapiens) в среде своего существования. В традиционной среде обитания значение этого числа для человека составляет около 150 связей³. Возможности действующих инфокоммуникационных сетей снимают жесткие ограничения на значение числа Данбара, и, как показывают недавние исследования, среднее количество устойчивых взаимосвязей индивида в сети Интернет может быть более шестисот⁴.

Особенностью первых десятилетий XXI века является мобилизация пользователей Интернета через самоусиливающийся контент социальных сетей. При этом, если объемы передаваемой информации становятся теоретически неограниченными, то время, затрачиваемое на создание, анализ, обработку и распространение информации, становится все более ограниченным ресурсом. Это явление ряд авторов определяют как «дефицит внимания», указывая, что за внимание пользователей идет интенсивная борьба крупных корпораций, распространителей рекламы и торговых сетей⁵.

В докладе Всемирного экономического форума в Давосе 2021 года отмечается, что современное общество становится все более поляризованным, разделяемым на сегменты увеличивающимися разрывами в цифровых технологиях, подвергается интенсивному воздействию дезинформации, что приводит к рискам появления «низших» классов в цифровом мире⁶.

Серьезное внимание проблеме дезинформации⁷ и манипулирования общественным мнением уделяет Европейская Комиссия⁸ (ЕС), которая призывает ведущие ин-

¹ Reisach U. The responsibility of social media in times of social and political manipulation // European Journal of Operational Research. 2021. Vol. 291. Issue 3. P. 906–917.

² Dunbar R. I. Neocortex size as a constraint on group size in primates // Journal of Human Evolution. 1992. Vol. 22 (6). P. 469–493.

³ Dunbar R. I. How many friends does one person need?: Dunbar's number and other evolutionary quirks. London: Faber and Faber, 2010. 302 p.

⁴ Dimitrova D., Wellman B. The Netlab Network // Advanced Methodologies and Technologies in Media and Communications. Mehdi Khosrow-Pour, D.B.A., 2019. P. 441–454.

⁵ Ryan C. D., Schaul A. J., Butner R., Swarthout J. T. Monetizing Disinformation in the attention economy: The case of genetically modified organisms (GMOs) // European Management Journal. 2020. Vol. 38. Issue 1. P. 7–18.

⁶ The Global Risks Report 2021. 16th Edition. URL: www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf. (дата обращения: 1.09.2022).

⁷ European Commission. Digitale Wirtschaft – Kommission strebt mehr Wettbewerb im Internet an (Digital Economy – Commission strives for more competition in the internet). Vertretung in Germany // Representation in Germany: EU-News no. 11/2020, 1–2. URL: https://ec.europa.eu/germany/sites/germany/files/docs/eu_nachrichten_11_2020web.pdf (дата обращения: 30.06.2020).

⁸ European Commission. European data strategy. Strategy priorities 2019–2024: A Europe fit for the digital age. Making the EU a role model for a society empowered by data Brussels, 19 February

тернет-платформы установить и реализовать последовательную и прозрачную модерацию информации в социальных сетях для обеспечения надежного устранения дезинформации¹.

В этой связи возникают проблемы определения достоверности информации, получаемой посредством сервисов Интернета и оценки влияния полученной информации на поведение экономических и социальных субъектов.

Основная часть

Современное состояние проблемы обнаружения дезинформации в инфокоммуникационных сетях

Сформировавшиеся в последнее десятилетие подходы к обнаружению дезинформации и недостоверных новостей в инфокоммуникационных сетях можно классифицировать следующим образом:

1. Методы, основанные на моделях распространения новостей в сети², представляющие процессы распространения сообщений³ в форме сложных графов⁴ с различными типами вершин⁵.

2. Методы, основанные на статистическом анализе текстов сообщений, исследующие статистику вхождений слов в корпус текстовых данных⁶, учитывающие совместное вхождение слов в тексты⁷, использующие линейный дискриминантный анализ⁸, использующие арифметические операции над векторами слов для представления синтаксических и семантических отношений⁹.

2020. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en. (дата обращения: 1.09.2022).

¹ European Commission (2019). Media Convergence and social media (Unit I.4): strategy-shaping Europe's digital future—Policy—tackling online disinformation. <https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation> (дата обращения: 30.06.2020).

² Hovy D. The enemy in your own camp: How well can we detect statistically-generated fake reviews—an adversarial study // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2016. Vol. 2. P. 351–356.

³ Potthast M., Kiesel J., Reinartz K., Bevendorff J., Stein B. Astylometric inquiry into hyperpartisan and fake news // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018. P. 231–240.

⁴ Kucharski A. Study epidemiology of fake news // Nature 540, 525 (2016). <https://doi.org/10.1038/540525a> (дата обращения: 1.09.2022).

⁵ Vosoughi S., Roy D., Aral S. The spread of true and false news online // Science. 2018. Vol. 359(6380). P. 1146–1151.

⁶ Kumar S., Shah N. False information on web and social media: A survey. arXiv preprint arXiv:1804.08559. 2018.

⁷ Fu X., Liu W., Xu Y., Cui L., 2017. Combine hownet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis // Neurocomputing. Vol. 241. P. 18–27.

⁸ Zhang Y., Lease M., Wallace B. Active Discriminative Text Representation Learning // Proceedings of the AAAI Conference on Artificial Intelligence, 2017 31(1). Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/10962> (дата обращения: 1.09.2022).

⁹ Caliskan A., Bryson J.J., Narayanan A. Semantics derived automatically from language corpora contain human-like biases // Science. 2017. Vol. 356. P. 183–186.

3. Методы, основанные на применении нейронных сетей и глубокого обучения. К этому классу относится наибольшее число публикаций, в которых описывается экспериментально подтвержденная эффективность обнаружения недостоверных новостей на различных массивах данных до уровня более 90 %.

Так, в работе¹ предложена архитектура глубокого полууправляемого обучения, содержащая три сверточные нейронные сети (СНН), первая из которых обучается извлекать из потока данных шаблоны недостоверных новостей. Выход этой сети подается одновременно на две СНН, одна из которых обучается на размеченных данных, а вторая – на неразмеченных данных. Формируются два вектора предсказания z_i и z_i' для одного и того же сообщения x_i . В процессе обучения минимизируется среднеквадратичная ошибка предсказания.

Эффективность данной модели оценивалась на двух эталонных наборах данных LIAR [paperswithcode.com/dataset/liar] и PHEME [pheme.eu]. LIAR включает в себя 12836 коротких сообщений, собранных с сайта PolitiFact2, посвященных различным темам (выборы, интервью, реклама, Twitter, Facebook и т. п.). Каждое сообщение трактуется как высказывание, помечаемое на множестве с шестью степенями истинности (истина, ложь, наполовину правда, наполовину ложь, неправда, в основном правда, едва правда). В модели использованы обучающее множество из 10269 высказываний, контрольное множество из 1284 высказываний и тестовое множество из 1283 высказываний.

Получаемая при использовании данной архитектуры точность выявления дезинформации сильно зависит от соотношения помеченных и непомеченных анализируемых сообщений и колеблется от 79,81 до 83,36 % при доле помеченных и непомеченных сообщений соответственно 1 и 30 %.

В работе² предложена сверточная нейронная сеть, настраиваемая с помощью алгоритмов глубокого обучения с гиперпараметрической оптимизацией (FNDNet – Fake News Detection Net). По утверждениям данных авторов, их модель FNDNet достигает точности 98,36 %, но точные размеры использованных наборов данных и их тематическая направленность не указаны.

В работе³ методы глубокого обучения применяются к сверточным нейронным сетям, использующим в качестве множества входных данных раскрашенные графы распространения информации в сети Интернет. В отличие от других подходов, использующих нейронные сети, авторы этой работы собирали информацию о распространении конкретных новостных сообщений, связывая истинностные оценки с конкретными URL-адресами источников и ретрансляторов сообщений. Новостные сообщения предварительно распределялись на два подмножества – истинные и ложные – на основе данных профессиональных некоммерческих организаций Snopes

¹ Dong X., Victor U., Chowdhury S., Qian L. Deep Two-path Semi-supervised Learning for Fake News Detection. 2019. arXiv preprint arXiv:1906.05659.

² Kaliyar R.K., Goswami A., Narang P., Sinha S., FNDNet – A Deep Convolutional Neural Network for Fake News Detection // Cognitive Systems Research. 2020. Vol. 61. Issue C. P. 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>.

³ Monti F., Frasca F., Eynard D., Mannion D., Bronstein M. M. Fake News Detection on Social Media using Geometric Deep Learning. 2019. arXiv preprint arXiv:1902.06673.

[<https://www.snopes.com/>], PoliFact2 [polifact.com], BuzzFeed3 [www.buzzfeed.com]. При этом отбрасывались сообщения, не имеющие однозначной оценки. Для отфильтрованных сообщений определялись связанные с ними URL-адреса, которые помечались как истинные или ложные в соответствии с оценкой рассматриваемого сообщения на их страницах. Качество этой оценки дополнительно проверялось экспертами. Для каждого новостного сообщения строились каскадные деревья их распространения на основе данных из Twitter.

В целом авторы данного подхода обработали 1084 сообщения, распространенных в Twitter в 2013–2018 годах. В процессах распространения сообщений было учтено 202375 уникальных пользователей, соответствующий граф информационных связей включал 2443996 ребер.

В данной модели использована четырехслойная СНН с двумя сверточными и двумя полносвязными слоями. Обучающее, тестирующее и контрольное множества содержали 667, 226 и 226 интернет-адресов соответственно с количеством истинных меток 83,26 % и ложных меток – 16,74 %. Обученная таким образом СНН показала точность обнаружения недостоверных сообщений от 88 до 92 %.

Данная модель интересна тем, что она непосредственно не связана с конкретным естественным языком, а ориентируется на процессы распространения информации в сети Интернет.

4. Модели, основанные на методах лексического, синтаксического и семантического анализа текстов сообщений. В работе¹ распространение дезинформации в социальных сетях рассматривается через призму явлений поляризации и сегрегации пользователей в среде Интернета².

Пользователи сетей склонны ограничивать сферу своего внимания немногими информационными источниками, формируя структуру своего новостного пространства. При этом в процессе обсуждения и дебатов в сетевом пространстве практически неизбежно возникает поляризация людей и противопоставление мнений. В статье³ предложена модель для быстрого определения поляризующего контента, который может служить почвой для последующего распространения дезинформации на платформах социальных сетей. Авторы разделяют источники информации на две категории: официальные (сообщения которых считаются достоверными) и неофициальные (сообщения которых в той или иной степени являются недостоверными). Из собранного массива сообщений обеих категорий извлекаются синтаксические и семантические характеристики текстов, такие как предмет обсуждения, участвующие лица, географические места, концепты, объекты, понятия. Для обсуждаемых новостей и комментариев анализируются настроения авторов сообщений и формируются

¹ Vicario M. D., Quattrociocchi W., Scala A., Zollo F. Polarization and fake news: Early warning of potential misinformation targets // ACM Transactions on the Web (TWEB). 2019. Vol. 13(2). P. 1–22.

² Zollo F., Bessi A., Del Vicario M., Scala A., Caldarelli G., Shekhtman L., Havlin S., Quattrociocchi W. Debunking in a world of tribes // PLOS ONE. 2017. Vol. 12. No. 7. P. 1–27.

³ Vicario M. D., Quattrociocchi W., Scala A., Zollo, F. Polarization and fake news: Early warning of potential misinformation targets // ACM Transactions on the Web (TWEB). 2019. Vol. 13(2). P. 1–22.

соответствующие оценки – от абсолютно отрицательных до абсолютно положительных.

На основе определенных интернет-ресурсов формируются интернет-сообщества пользователей, представляющих собой малые социальные группы, весьма чувствительные к информации, противоречащей коллективным базовым убеждениям¹. Воздействующая на такие сообщества через сервисы Интернета дезинформация при дефиците достоверной информации провоцирует панические, тревожные эмоциональные состояния, снижающие рефлексивность, запускает защитные групповые механизмы, поддерживающие позитивные оценки группы ее членами.

Воздействия через Интернет, формируемые различными группами в социальных сетях и сообществах, могут носить как выраженный, так и явно не выраженный деструктивный и экстремистский характер, направленный на искусственное конструирование несуществующих проблем, провоцирование общественного ответа в форме интенсивных коллективных переживаний и, наоборот, нацеленный на нивелирование реальных проблем, снижение остроты социальной реакции².

Развитие интернет-сервисов (социальных сетей, сетевых сообществ, мессенджеров и т. п.) породило проблему «астротурфинга»³ (массового найма блоггеров и троллей⁴, разворачивающих в социальных сетях определенным образом ориентированную информационную деятельность⁵, имитируя общественное мнение в государственных или корпоративных целях⁶).

Выявлен ряд информационно-технических и социально-психологических эффектов, способствующих распространению дезинформации через астротурфинг: множественность источников, повторяющих данную информацию⁷, неопределенность социальной оценки⁸, воспринимаемость сходства и подобие мотивации пользователей⁹.

¹ Нестик Т. А. Отношение к будущему в российских управленческих командах: лидерское видение и корпоративный форсайт // Экономические стратегии. 2014. № 2. С. 134–141.

² Емельянова Т. П. Социальные представления: История, теория и эмпирические исследования. Москва: Институт психологии РАН, 2016. 476 с.

³ Hancock J. T. The future of lying. Technology, Education & Design (TED). Winnipeg, MN, 2012. URL: http://www.ted.com/talks/jeff_hancock_3_types_of_digital_lies. (дата обращения: 1.09.2022).

⁴ Hancock J. T., Woodworth M. T., Goorha S. See no evil: The effect of communication medium and motivation on deception detection // Group Decision and Negotiation. 2010. Vol. 19. P. 327–343.

⁵ Weedon J., Nuland W., Stamos A. Information Operations and Facebook. April 27, 2017. Facebook Inc., 2017.

⁶ Виловатых А.В. Социальные движения в эпоху постправды // Проблемы национальной стратегии. 2020. №1 (58). С. 90–104.

⁷ Zhang J., Carpenter D., Ko M. Online astroturfing: A theoretical perspective // Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15–17, 2013. URL: https://www.researchgate.net/profile/Darrell_Carpenter/publication/286729041 (дата обращения: 1.09.2022).

⁸ Докука С. Практики использования онлайн-социальных сетей // Социологические исследования. 2014. № 1. С. 138–145.

⁹ Kovic M., Rauchfleisch A., Sele M., Caspar C. Digital astroturfing in politics: Definition, typology, and countermeasures // Studies in Communication Sciences. 2018. Vol. 18.1. P. 69–85.

Различные аргументы от разных пользователей воспринимаются индивидом как более весомые, чем один и тот же аргумент от разных источников или разные аргументы от одного пользователя.

Интернет-пользователи более подвержены информационному влиянию при обсуждении вопросов, по которым еще не сложилось однозначное общественное мнение, более восприимчивы к мнению субъектов, подобных себе, с большой вероятностью у них нет высокой мотивации детально разбираться в соответствующем вопросе.

В последние годы выявлена проблема устойчивости внедренной в сознание интернет-пользователей дезинформации, невосприимчивость индивидов к опровержению недостоверной информации в тех случаях, когда эта недостоверная информация соответствует сложившимся установкам человека.

Значительный объем современного интернет-пространства занимают социальные сети, охватывающие в настоящее время миллиарды пользователей, формирующие новые специфические социальные взаимоотношения, субъектами которых выступают отдельные люди, группы людей, сообщества. Общение в социальных сетях можно рассматривать как компьютерно-опосредованную коммуникацию, в ходе которой распространяется достоверная, искаженная и недостоверная информация, оказывающая воздействие на различные группы интернет-пользователей.

В ряде работ исследовалось коммуникативное поведение в социальных сетях¹ с изучением различных семантико-прагматических факторов коммуникации²: тема коммуникации, микро-/макроконтекст, ситуация общения, тотальность коммуникации, регистр коммуникации, коммуникационные роли, темпоральный, локальный, культурный аспекты³.

В этой связи представляется целесообразным исследование лингво-семантических особенностей достоверной, искаженной и недостоверной информации, циркулирующей в социальных сетях Интернета через оценку диапазона и качества речевых единиц, коммуникативных паттернов, тем и смыслового содержания текстовых сообщений.

Результаты проведенного анализа показывают, что в настоящее время сформировался значительный спектр методов и моделей распространения информации в сети Интернет и оценки достоверности получаемых пользователями сообщений. Разработанные подходы имеют свои достоинства и преимущества относительно других типов моделей, но, вместе с тем, характеризуются и рядом существенных недостатков. Так, методы, основанные на моделях распространения новостей в сети, представляющие процессы распространения сообщений в форме сложных графов с различными

¹ Рипшинен Т. О., Слободская Е. Р. Взаимосвязи личностных особенностей подростков с повседневным использованием компьютера // Психологический журнал. 2014. Т. 35. № 4. С. 30–37.

² Смолина Е. Г. Понятие «сети» как новый взгляд на устройство общества // Научный вестник Волгоградского филиала РАНХиГС. Серия: Политология и социология. 2015. № 2. С. 23–28.

³ Сушков И. Р., Козлова Н. С. Интернет-активность как проявление потребности личности в коллективном субъекте // Психологический журнал. 2015. Т. 36. № 5. С. 75–83.

типами вершин, не учитывают содержание передаваемых сообщений и не чувствительны к семантике транслируемых текстов. Методы, основанные на статистическом анализе текстов сообщений, не учитывают семантику исследуемых текстов и онтологию рассматриваемых предметных областей, а также требуют накопления статистических данных и массивов текстов на заданную тематику. Методы, основанные на применении нейронных сетей и глубокого обучения, предполагают обучение на эталонных наборах, требуют формирования или наличия обучающих, контрольных и тестовых множеств сообщений или высказываний, что затрудняет их применение в реальном масштабе времени при необходимости анализа ситуативно появляющихся новых сообщений.

Модели, основанные на методах лексического, синтаксического и семантического анализа текстов сообщений, не учитывают временной аспект возникновения сообщений о событии, не направлены на точное установление первоисточника информации (или дезинформации) и в основном ориентированы на обработку английского языка. Таким образом, сформировались условия для исследования теоретических методов и моделей, разработки алгоритмов и программных средств анализа сообщений и выявления дезинформации в сети Интернет, учитывающих содержание и смысл обрабатываемых текстов, выявляющих временную динамику распространения дезинформации, устанавливающих первоисточники возникновения дезинформации, при этом базирующихся на онтологиях анализируемых предметных областей.

В соответствии с концепцией Д. И. Дубровского¹, субъектами процесса распространения дезинформации могут быть отдельные люди (**SI**); коллективы людей, связанные между собой различными отношениями (**SK**) – родственными, деловыми, идеологическими, религиозными и т. п.; большие социальные группы, классы, народы, образующие массового субъекта (**SM**); социальные институты общества, формирующие институциональные субъекты (**ST**) (рис. 1).



Рис. 1. Субъекты процесса обмана

¹ Дубровский Д. И. Обман. Философско-психологический анализ. Москва: Конон+, 2010. 336 с.

Субъекты данных категорий рассматриваются нами как интеллектуальные агенты, которым следует приписать следующие возможные внутренние состояния: (А) – осуществления дезинформирования, т. е. передачи заведомо ложной информации получателю, (В) – воздействия обмана, или нахождения в состоянии приема дезинформации, (С) – жертвы – получатель дезинформации поверил в ее истинность, (D) – неудавшейся жертвы (получатель не поверил дезинформации и не считает полученное сообщение истинным).

Между субъектами процесса обмана возникают бинарные отношения, такие как: R_{AB} (субъект S_i , находящийся в состоянии А обманывает в настоящее время субъекта S_j , находящегося в состоянии В); R_{AC} (субъект S_i , находящийся в состоянии А, обманул к настоящему времени субъекта S_j , перешедшего к данному моменту времени в состояние С); R_{AD} (субъект S_i , находящийся в состоянии А, не смог обмануть к настоящему времени субъекта S_j , и данный субъект перешел к данному моменту времени в состояние D).

Если типы субъектов SI, SK, SM, ST связать с вершинами плоского ориентированного графа, а отношения R_{AB}, R_{AC}, R_{AD} приписать к дугам, связывающим возможные пары типов субъектов, то можно построить полносвязный ориентированный граф взаимосвязи субъектов процесса обмана G_{des} , в котором вершины SI, SK, SM, ST имеют петли, так как возможен самообман субъектов всех типов (рис. 2).

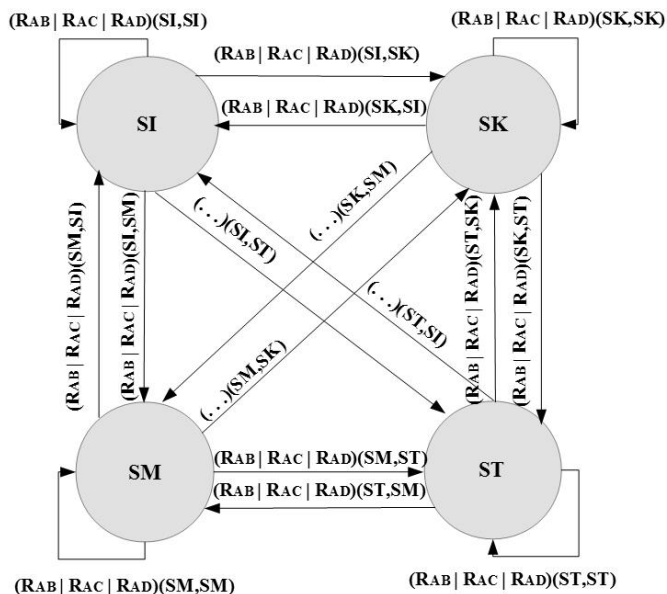


Рис. 2. Граф обмена G_{des}

Формула $(R_{AB} | R_{AC} | R_{AD})(S_i, S_j)$ показывает, что в данный момент времени может выполняться только одно из отношений R_{AB}, R_{AC}, R_{AD} между субъектами S_i и S_j (символ "|" означает альтернативу).

В графе G_{des} возможны следующие циклы по вершинам и дугам: четыре цикла, включающие по одной вершине, шесть циклов, содержащих по две вершины, четыре цикла по три вершины и один цикл, охватывающий все четыре вершины. При дальнейшем исследовании структуры графа G_{des} необходимо учитывать возможные циклы при переходах по дугам, нагруженным различными отношениями из множества $\{R_{AB}, R_{AC}, R_{AD}\}$, образующими сочетания в двух- и трехвершинных циклах и в одном цикле из четырех вершин.

Поскольку принимающий субъект S_j может относиться к информации передающего субъекта S_i с разной степенью доверия (от абсолютного доверия до полного недоверия), то предполагается дальнейшее исследование существования возможных циклов дезинформации в модальной логике знания и веры¹.

Модель агента процесса распространения дезинформации

В широком контексте субъекты обозначенных выше классов SI, SK, SM, ST могут рассматриваться как вероятностные автоматы, реализующие некоторое стохастическое поведение в процессе приема и обработки информации, получаемой от внешних источников. Граф состояний и переходов такого вероятностного автомата для обобщенного субъекта G_{Sdes} представлен на рис. 3 и содержит следующие состояния и сигналы.

¹ Бжанишвили М. Н. Логика модальностей знания и мнения. Москва: КомКнига, 2007. 288 с.

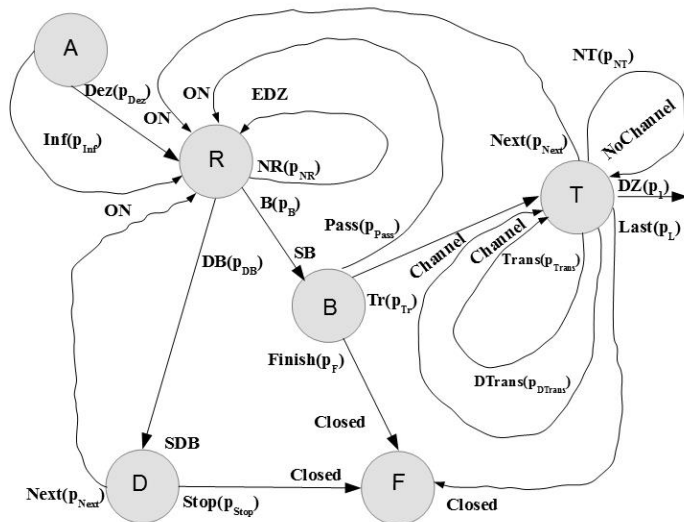


Рис. 3. Граф состояний и переходов вероятностного автомата агента процесса распространения дезинформации

A – источник дезинформации, отправляющий сообщения дезинформации Dez с вероятностью p_{dez} и истинные сообщения Inf с вероятностью p_{inf} .

R – состояние приема информации, из которого возможны три перехода: по сигналу NR (отказ от приема, отбраковка входных сообщений) с вероятностью p_{NR} и выдачей сигнала EDZ на уничтожение пришедшего отбракованного сообщения, по сигналу $B(belief)$ – принятие входного сообщения как достоверного и запись его в базу знаний (БЗ) по сигналу SB с вероятностью p_B , по сигналу $DB(not\ belief)$ – неверие в истинность входного сообщения (полагание его ложным) с вероятностью p_{DB} и формирование сигнала SDB для записи в БЗ информации \overline{DZ} и перехода в состояние D . B – внутреннее состояние доверия (веры) в принятое сообщение Dez , из которого возможны следующие переходы: по сигналу $Pass(Passive\ Belief)$ с вероятностью p_{Pass} , что означает пассивную веру субъекта в сообщении (информация принята, но далее не транслируется); по сигналу TR с вероятностью p_{TR} информация передается к состоянию трансляции T ; по сигналу $F(finish)$ с вероятностью p_F автомат переходит в заключительное тупиковое состояние F , из которого нет выхода в текущем процессе обработки сообщений.

T – внутреннее состояние передачи информации следующим субъектам информационного процесса, из которого возможны четыре перехода: по сигналу $Next$ с вероятностью p_{Next} переход в состояние R с выдачей сообщения ON , означающего включение ожидания следующего сообщения из информационного потока от источника A ; по сигналу $Trans$ с вероятностью p_{Trans} происходит передача сообщения DZ в канал передачи информации, т. е. во внешнюю относительно данного субъекта S_i информационную среду; по сигналу $DTrans$ с вероятностью p_{DTrans} осуществляется повторная передача сообщения DZ , которая может в общем случае выполняться мно-

гократно; по сигналу $Last$ с вероятностью p_L передача информации во внешнюю среду прекращается и происходит переход в финальное состояние F ; по сигналу NT , означающему невозможность передачи сообщения по каким-либо причинам, происходит возврат в состояние T с вероятностью p_{NT} .

D – внутреннее состояние недоверия к принятому сообщению DZ , из которого возможны два перехода: по сигналу $Stop$ с вероятностью p_{Stop} переход в финальное состояние F ; по сигналу $Next$ с вероятностью p_{Next} – возврат в состояние R с разрешением приема следующих сообщений сигналом ON .

Исходя из структуры графа состояний и переходов G_{Sdess} , получаются следующие соотношения для вероятностей возможных переходов:

$$\begin{aligned} p_{NR} + p_B + p_{DB} &= 1 \\ p_{Pass} + p_{Tr} + p_F &= 1 \\ p_{NT} + p_L + p_{DTrans} + p_{Trans} + p_{Next} &= 1 \\ p_{Next} + p_{Stop} &= 1. \end{aligned}$$

При этом вероятность дальнейшего распространения сообщений как достоверных, так и недостоверных, определяется как

$$p_1 = p_{Trans} + p_{DTrans} = 1 - (p_{NT} + p_L + p_{Next}).$$

Субъекты введенных нами классов SI , SK , SM , ST являются по существу интеллектуальными агентами в трактовке¹, поведение которых может описываться некоторым набором моделей. Такие субъекты образуют пространственно-временную сеть, распространение дезинформации в которой мы будем называть процессом распространения дезинформации (ПРД), а агентов, включенных в данную сеть и участвующих в ПРД, – агентами ПРД. В рамках стохастического подхода к ПРД агенты могут рассматриваться как вероятностные автоматы (рис. 3).

Структуру сети распространения дезинформации представляем ориентированным связным графом, который называется далее графом информационных связей (ГИС-ПРД) процесса распространения дезинформации, в котором вершинами являются агенты ПРД (SI_i , SK_i , SM_i , ST_i), в общем случае обозначаемые SC_i (под C понимается I , K , M , T), а дуги соответствуют наличию и направлению передачи информации от агента SC_i к агенту SC_j :

$$GIS = G(SC_i | i = 1 \dots n, E),$$

где E – множество дуг информационных связей (e_1, e_2, \dots, e_m) .

¹ Швецов А. Н. Агентно-ориентированные системы: основные модели. Вологда: Вологодский государственный университет, 2012. 190 с.

Графу информационных связей будет соответствовать матрица смежности, называемая далее матрицей информационных связей (МИС): $MIS = \|e_{ij}\|$, где $e_{ij} = 1$ соответствует наличию связи по передаче информации $SC_i \rightarrow SC_j$, $e_{ij} = 0$ означает отсутствие передачи информации от SC_i к агенту SC_j .

$$MIS = \begin{vmatrix} 0 & e_{12} & \dots & e_{1n} \\ e_{21} & 0 & \dots & e_{2n} \\ \dots & \dots & 0 & \dots \\ e_{n1} & \dots & \dots & 0 \end{vmatrix}.$$

Особенностью информационно-телекоммуникационных сетей (например, социальных сетей в Интернете) является динамический характер взаимосвязей $SC_i \rightarrow SC_j$, поскольку каждый агент социальной сети может передавать или не передавать сообщения разному количеству других участников сети. Каждому сообщению Dez может соответствовать своя геометрия ГИС-ПРД, поэтому целесообразно исследовать процессы распространения дезинформации применительно к каждому конкретному сообщению Dez , и только потом искать более общие закономерности.

Структура ГИС-ПРД определяет потенциально возможное множество путей распространения дезинформации в рассматриваемой сети и оценки вероятности получения дезинформации конкретным агентом инфокоммуникационной сети.

Выводы

Предложенные модели агентов процессов распространения дезинформации, графа и матрицы информационных связей позволяют формализовать основные элементы процесса распространения дезинформации в рамках агентно-ориентированного подхода к анализу информационно-телекоммуникационных сетей.

Дальнейшие исследования предполагают детализацию процессов распространения дезинформации с учетом начального множества источников дезинформации, степени связности вершин в графе информационных связей, ограничений на время и скорость распространения информации по каналам инфокоммуникационной сети.

В настоящее время разрабатывается имитационная агентно-ориентированная модель процессов распространения дезинформации в среде мультиагентного моделирования NetLogo 6.2.2¹, которая позволит оценить адекватность предложенного подхода к проблеме распространения дезинформации в инфокоммуникационных сетях.

¹ Wilensky U., Rand W. An introduction to agent-based modeling: Modeling natural, social and engineered complex systems with NetLogo. Cambridge: MIT Press, 2015. 482 p.

Список источников

- Бежанишвили М. Н. Логика модальностей знания и мнения. Москва: КомКнига, 2007. 288 с.
- Виловатых А. В. Социальные движения в эпоху постправды // Проблемы национальной стратегии. 2020, №1 (58). С. 90–104.
- Докука С. Практики использования онлайн-социальных сетей // Социологические исследования. 2014. № 1. С. 138–145.
- Дубровский Д.И. Обман. Философско-психологический анализ. Москва: Конон+, 2010. 336 с.
- Емельянова Т. П. Социальные представления: История, теория и эмпирические исследования. Москва: Изд-во «Институт психологии РАН», 2016. 476 с.
- История информатики и философия информационной реальности / под редакцией Р. М. Юсупова, В. П. Котенко. Москва: Академический проект, 2007. 430 с.
- Колин К. К. Философские проблемы информатики. Москва: БИНОМ, 2010. 264 с.
- Нестик Т. А. Отношение к будущему в российских управленческих командах: лидерское видение и корпоративный форсайт // Экономические стратегии. 2014. № 2. С. 134–141.
- Риппинен Т. О., Слободская Е. Р. Взаимосвязи личностных особенностей подростков с повседневным использованием компьютера // Психологический журнал. 2014. Т. 35. № 4. С. 30–37.
- Смолина Е.Г. Понятие «сети» как новый взгляд на устройство общества // Научный вестник Волгоградского филиала РАНХиГС. Серия: Политология и социология. 2015. № 2. С. 23–28.
- Сушков И. Р., Козлова Н. С. Интернет-активность как проявление потребности личности в коллективном субъекте // Психологический журнал. 2015. Т. 36. № 5. С. 75–83.
- Швецов А. Н. Агентно-ориентированные системы: основные модели. Вологда: Вологодский государственный университет, 2012. 190 с.
- Caliskan A., Bryson J. J., Narayanan A. Semantics derived automatically from language corpora contain human-like biases // Science. 2017. Vol. 356. P. 183–186.
- Colic-Peisker V., Flitney A. The Age of Post-Rationality: Limits of economic reasoning in the 21st century. Singapore: Palgrave MacMillan, 2018. 253 p.
- Digital 2022 July Global Statshot Report. URL: <https://datareportal.com/reports/digital-2022-july-global-statshot?rq=Digital%202022/> (дата обращения 01.09.2022).
- Dimitrova D., Wellman B. The Netlab Network // Advanced Methodologies and Technologies in Media and Communications. Mehdi Khosrow-Pour, D.B.A., 2019. P. 441–454.
- Dong X., Victor U., Chowdhury S., Qian L. Deep Two-path Semi-supervised Learning for Fake News Detection. 2019. arXiv preprint arXiv:1906.05659.
- Dunbar R. I. How many friends does one person need?: Dunbar's number and other evolutionary quirks. London: Faber and Faber, 2010. 302 p.
- Dunbar R. I. Neocortex size as a constraint on group size in primates // Journal of Human Evolution. 1992. Vol. 22(6). P. 469–493.
- European Commission (2019). Media Convergence and social media (Unit I.4): strategy–shaping Europe’s digital future–Policy–tackling online disinformation. <https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation> (дата обращения: 30.06.2020).
- European Commission. Digitale Wirtschaft – Kommission strebt mehr Wettbewerb im Internet an (Digital Economy – Commission strives for more competition in the internet). Vertretung in Germany. In: Representation in Germany: EU-News no. 11/2020, 1-2. https://ec.europa.eu/germany/sites/germany/files/docs/eu_nachrichten_11_2020web.pdf (дата обращения: 30.06.2020).

European Commission. European data strategy. Strategy priorities 2019–2024: A Europe fit for the digital age. Making the EU a role model for a society empowered by data Brussels, 19 February 2020 https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

Fallis D. The Epistemic Threat of Deepfakes // *Philosophy&Technology*. 2021. Vol. 34. P. 623–643.

Fu X., Liu W., Xu Y., Cui L., 2017. Combine hownet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis // *Neurocomputing*. Vol. 241. P. 18–27.

Hancock J. T. The future of lying. *Technology, Education & Design (TED)*. Winnipeg, MN, 2012. URL: http://www.ted.com/talks/jeff_hancock_3_types_of_digital_lies (дата обращения: 1.10.2022)

Hancock J. T., Woodworth M. T., Goorha S. See no evil: The effect of communication medium and motivation on deception detection // *Group Decision and Negotiation*. 2010. Vol. 19. P. 327–343.

Hovy D. The enemy in your own camp: How well can we detect statistically-generated fake reviews—an adversarial study // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2016. Vol. 2. P. 351–356.

Kaliyar R.K., Goswami A., Narang P., Sinha S., FNDNet – A Deep Convolutional Neural Network for Fake News Detection // *Cognitive Systems Research*. 2020. Vol. 61. Issue C. P. 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>.

Kovic M., Rauchfleisch A., Sele M., Caspar C. Digital astroturfing in politics: Definition, typology, and countermeasures // *Studies in Communication Sciences*. 2018. Vol. 18.1. P. 69–85.

Kucharski A. Study epidemiology of fake news // *Nature* 540, 525 (2016). <https://doi.org/10.1038/540525a>

Kumar S., Shah N. False information on web and social media: A survey. arXiv preprint arXiv:1804.08559. 2018.

Kumar S., West R., Lescovec J. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes // *WWW '16: Proceedings of the 25th International Conference on World Wide Web April 2016*. P. 591–602. URL: <https://doi.org/10.1145/2872427.2883085>.

Monti F., Frasca F., Eynard D., Mannion D., Bronstein M. M. Fake News Detection on Social Media using Geometric Deep Learning. 2019. arXiv preprint arXiv:1902.06673.

Potthast M., Kiesel J., Reinartz K., Bevendorff J., Stein B. Astylometric inquiry into hyperpartisan and fake news // *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018. P. 231–240.

Reisach U. The responsibility of social media in times of social and political manipulation // *European Journal of Operational Research*. 2021. Vol. 291. Issue 3. P. 906–917.

Ryan C. D., Schaul A. J., Butner R., Swarthout J. T. Monetizing Disinformation in the attention economy: The case of genetically modified organisms (GMOs) // *European Management Journal*. 2020. Vol. 38. Issue 1. P. 7–18.

Soe S. O. A Unified Account of Information, Misinformation, and mation / In: *Synthese – An international journal for epistemology, methodology and philosophy of science*. 2021. Vol. 198. P. 5929–5949.

The Global Risks Report 2021. 16th Edition. URL: www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf (дата обращения: 1.09.2022)

Vicario M. D., Quattrociochi W., Scala A., Zollo, F. Polarization and fake news: Early warning of potential misinformation targets // *ACM Transactions on the Web (TWEB)*. 2019. Vol. 13(2). P. 1–22.

Vosoughi S., Roy D., Aral S. The spread of true and false news online // *Science*. 2018. Vol. 359(6380). P. 1146–1151.

Weedon J., Nuland W., Stamos A. Information Operations and Facebook. April 27, 2017. Facebook Inc., 2017.

Wilensky U., Rand W. An introduction to agent-based modeling: Modeling natural, social and engineered complex systems with NetLogo. Cambridge: MIT Press, 2015. 482 p.

Zhang J., Carpenter D., Ko M. Online astroturfing: A theoretical perspective // Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15–17, 2013. URL: https://www.researchgate.net/profile/Darrell_Carpenter/publication/286729041 (дата обращения: 1.09.2022)

Zhang Y., Lease M., Wallace B. Active Discriminative Text Representation Learning // Proceedings of the AAAI Conference on Artificial Intelligence, 2017 31(1). Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/10962> (дата обращения: 1.08.2022).

Zollo F., Bessi A., Del Vicario M., Scala A., Caldarelli G., Shekhtman L., Havlin S., Quattrociocchi W. Debunking in a world of tribes // PLOS ONE. 2017. Vol. 12. No. 7. P. 1–27.

References

Bezhanishvili M. N. *Logika modal'nostei znaniia i mneniia* [Logic of modalities of knowledge and opinion]. Moscow: KomKniga, 2007. 288 p.

Vilovatykh A. V. Sotsial'nye dvizheniia v epokhu postpravdy [Social movements in the age of post-truth]. *Problemy natsional'noi strategii* [Problems of national strategy], 2020, no. 1 (58), pp. 90–104.

Dokuka S. Praktiki ispol'zovaniia onlainovykh sotsial'nykh setei [Practices of using online social networks]. *Sotsiologicheskie issledovaniia* [Sociological studies], 2014, no. 1, pp. 138–145.

Dubrovskii D. I. *Obman. Filosofsko-psikhologicheskii analiz* [Deception. Philosophical and psychological analysis]. Moscow: Konon+, 2010. 336 p.

Emel'ianova T. P. *Sotsial'nye predstavleniia: Istoriia, teoriia i empiricheskie issledovaniia* [Social representations: history, theory and empirical research]. Moscow: Institut psikhologii RAN, 2016. 476 p.

Istoriia informatiki i filosofiiia informatsionnoi real'nosti [History of computer science and philosophy of information reality; ed. by P. M. Iusupov, V. P. Kotenko]. Moscow: Akademicheskii proekt, 2007. 430 p.

Kolin K. K. *Filosofskie problemy informatiki* [Philosophical problems of informatics]. Moscow: BINOM, 2010. 264 p.

Nestik T. A. Otnoshenie k budushchemu v rossiiskikh upravlencheskikh komandakh: liderskoe videnie i korporativnyi forsait [Attitude towards the future in Russian management teams: leadership vision and corporate foresight]. *Ekonomicheskie strategii* [Economic Strategies], 2014, no. 2, pp. 134–141.

Rippen T. O., Slobodskaja E. R. Vzaimosviazi lichnostnykh osobennostei podrostkov s povsednevnyim ispol'zovaniem komp'iutera [Relationship of adolescents' personal peculiarities with everyday use of computer]. *Psikhologicheskii zhurnal* [Psychological journal], 2014, Vol. 35, no. 4, pp. 30–37.

Smolina E.G. Poniatie "seti" kak novyi vzgliad na ustroistvo obshchestva [The concept of "network" as a new look at the structure of society]. *Nauchnyi vestnik Volgogradskogo filiala RANKhIGS. Seriya: Politologiya i sotsiologiya* [Scientific Bulletin of the Volgograd branch of the RANEPА. Series: Political science and sociology], 2015, no. 2, pp. 23–28.

Sushkov I. R., Kozlova N. S. Internet-aktivnost' kak proiavlenie potrebnosti lichnosti v kollektivnom sub"ekte [Internet activity as a manifestation of person's need for collective subject]. *Psikhologicheskii zhurnal* [Psychological journal], 2015, Vol. 36, no. 5, pp. 75–83.

- Shvetsov A. N. *Agentno-orientirovannye sistemy: osnovnye modeli* [Agent-oriented systems: basic models]. Vologda: Vologodskii gosudarstvennyi universitet, 2012. 190 p.
- Caliskan A., Bryson J. J., Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*, 2017, Vol. 356, pp. 183–186.
- Colic-Peisker V., Flitney A. *The Age of Post-Rationality: Limits of economic reasoning in the 21st century*. Singapore: Palgrave MacMillan, 2018. 253 p.
- Digital 2022 July Global Statshot Report*. Available at: <https://datareportal.com/reports/digital-2022-july-global-statshot?rq=Digital%202022/> (accessed: 01.09.2022).
- Dimitrova D., Wellman B. The Netlab Network. *Advanced Methodologies and Technologies in Media and Communications*. Mehdi Khosrow-Pour, D.B.A., 2019, pp. 441–454.
- Dong X., Victor U., Chowdhury S., Qian L. *Deep Two-path Semi-supervised Learning for Fake News Detection*. 2019. arXiv preprint arXiv:1906.05659.
- Dunbar R. I. *How many friends does one person need?: Dunbar's number and other evolutionary quirks*. London: Faber and Faber, 2010. 302 p.
- Dunbar R. I. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 1992, Vol. 22(6), pp. 469–493.
- European Commission (2019). *Media Convergence and social media (Unit I.4): strategy–shaping Europe’s digital future–Policy–tackling online disinformation*. Available at: <https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation> (accessed: 30.06.2020).
- European Commission. Digitale Wirtschaft – Kommission strebt mehr Wettbewerb im Internet an (Digital Economy – Commission strives for more competition in the internet). Vertretung in Germany. *Representation in Germany: EU-News* no.11/2020, 1-2. Available at: https://ec.europa.eu/germany/sites/germany/files/docs/eu_nachrichten_11_2020web.pdf (accessed: 30.06.2020).
- European Commission. *European data strategy. Strategy priorities 2019–2024: A Europe fit for the digital age. Making the EU a role model for a society empowered by data Brussels, 19 February 2020*. Available at: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en. (accessed: 30.06.2022).
- Fallis D. The Epistemic Threat of Deepfakes. *Philosophy&Technology*, 2021, Vol. 34, pp. 623–643.
- Fu X., Liu W., Xu Y., Cui L., 2017. Combine hownet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. *Neurocomputing*, Vol. 241, pp. 18–27.
- Hancock J. T. *The future of lying. Technology, Education & Design (TED)*. Winnipeg, MN, 2012. Available at: http://www.ted.com/talks/jeff_hancock_3_types_of_digital_lies. (accessed: 1.09.2022)
- Hancock J. T., Woodworth M. T., Goorha S. See no evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation*, 2010, Vol. 19, pp. 327–343.
- Hovy D. The enemy in your own camp: How well can we detect statistically-generated fake reviews—an adversarial study. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), 2016, Vol. 2, pp. 351–356.
- Kaliyar R.K., Goswami A., Narang P., Sinha S., FNDNet – A Deep Convolutional Neural Network for Fake News Detection. *Cognitive Systems Research*, 2020, Vol. 61, Issue C, pp. 32–44. Available at: <https://doi.org/10.1016/j.cogsys.2019.12.005>.
- Kovic M., Rauchfleisch A., Sele M., Caspar C. Digital astroturfing in politics: Definition, typology, and countermeasures. *Studies in Communication Sciences*, 2018, Vol. 18.1, pp. 69–85.
- Kucharski A. *Study epidemiology of fake news*. Nature 540, 525 (2016). Available at: <https://doi.org/10.1038/540525a>
- Kumar S., Shah N. *False information on web and social media: A survey*. arXiv preprint arXiv:1804.08559. 2018.

Kumar S., West R., Lescovec J. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. *WWW '16: Proceedings of the 25th International Conference on World Wide Web April 2016*. pp. 591–602. Available at: <https://doi.org/10.1145/2872427.2883085>.

Monti F., Frasca F., Eynard D., Mannion D., Bronstein M. M. *Fake News Detection on Social Media using Geometric Deep Learning*, 2019. arXiv preprint arXiv:1902.06673.

Potthast M., Kiesel J., Reinartz K., Bevendorff J., Stein B. Astylometric inquiry into hyperpartisan and fake news. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 231–240.

Reisach U. The responsibility of social media in times of social and political manipulation. *European Journal of Operational Research*, 2021, Vol. 291, Issue 3, pp. 906–917.

Ryan C. D., Schaul A. J., Butner R., Swarthout J. T. Monetizing Disinformation in the attention economy: The case of genetically modified organisms (GMOs). *European Management Journal*, 2020, Vol. 38, Issue 1, pp. 7–18.

Soe S. O. A Unified Account of Information, Misinformation, and Disinformation. *Synthese – An international journal for epistemology, methodology and philosophy of science*, 2021, Vol. 198, pp. 5929–5949.

The Global Risks Report 2021. 16th Edition. Available at: www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf. (accessed: 1.09.2022).

Vicario M. D., Quattrociochi W., Scala A., Zollo F. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 2019, Vol. 13(2), pp. 1–22.

Vosoughi S., Roy D., Aral S. The spread of true and false news online. *Science*, 2018, Vol. 359(6380), pp. 1146–1151.

Weedon J., Nuland W., Stamos A. *Information Operations and Facebook*. April 27, 2017, Facebook Inc, 2017.

Wilensky U., Rand W. *An introduction to agent-based modeling: Modeling natural, social and engineered complex systems with NetLogo*. Cambridge: MIT Press, 2015. 482 p.

Zhang J., Carpenter D., Ko M. *Online astroturfing: A theoretical perspective, Proceedings of the Nineteenth Americas Conference on Information Systems*. Chicago, Illinois, August 15–17, 2013. Available at: https://www.researchgate.net/profile/Darrell_Carpenter/publication/286729041. (accessed: 1.09.2022).

Zhang Y., Lease M., Wallace B. *Active Discriminative Text Representation Learning Proceedings of the AAAI Conference on Artificial Intelligence*, 2017 31(1). Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/10962> (accessed: 21.08.2022).

Zollo F., Bessi A., Del Vicario M., Scala A., Caldarelli G., Shekhtman L., Havlin S., Quattrociochi W. Debunking in a world of tribes, *PLOS ONE*, 2017, Vol. 12, no. 7. Pp. 1–27.

Сведения об авторах

Анатолий Николаевич Швецов – доктор технических наук, профессор; <https://orcid.org/0000-0002-7021-5184>, smithv@mail.ru, Вологодский государственный университет (д. 15, ул. Ленина, 160000 Вологда, Россия); **Anatolii N. Shvetsov** – Doctor of Technical Sciences, Professor; <https://orcid.org/0000-0002-7021-5184>, smithv@mail.ru, Vologda State University (15, ul. Lenina, 160000 Vologda, Russia).

Сергей Владимирович Дианов – кандидат технических наук; <https://orcid.org/0000-0001-8297-8077>, dianov,sv@mail.ru, Вологодский государственный университет (д. 15, ул. Ленина, 160000 Вологда, Россия); **Sergey V. Dianov** – Candidate of Technical Sciences;

<https://orcid.org/0000-0001-8297-8077>, dianov.sv@mail.ru, Vologda State University (15, ul. Lenina, 160000 Vologda, Russia).

Попова Анастасия Александровна – студент, afitskevich@bk.ru, Вологодский государственный университет (д. 15, ул. Ленина, 160000 Вологда, Россия); **Anastasia A. Popova** – student, afitskevich@bk.ru, Vologda State University (15, ul. Lenina, 160000 Vologda, Russia).

Кузнецов Дмитрий Олегович – студент, awesomeshout@yandex.ru, Вологодский государственный университет (д. 15, ул. Ленина, 160000 Вологда, Россия); **Dmitry O. Kuznetsov** – student, awesomeshout@yandex.ru, Vologda State University (15, ul. Lenina, 160000 Vologda, Russia).

Заявленный вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests,

Статья поступила в редакцию 01.10.2022; одобрена после рецензирования 20.10.2022; принята к публикации 14.11.2022.

The article was submitted 01.10.2022; Approved after reviewing 20.10.2022; Accepted for publication 14.11.2022.