Information Disorder Detection Applying a Hybrid NLP and Knowledge Graph-based Solution Approach

Mark Nicolas Gruensteidl

Institute for Complex Networks, Vienna University of Economics and Business, Vienna, Austria

Abstract

Extensive consumption of news and rapid communication flows on the web, especially via online content sharing platforms, leads to economic and societal harm caused by information disorders. These are commonly known as fake news and represent a threat to democratic processes and societies. While artificial intelligence-based models can detect information disorder types, current algorithms are not transparent, explainable, trust-building, and domain-specific enough. Therefore, the aim of this work is to advance the state-of-the-art in terms of information disorder detection and mitigation by combining explainable artificial intelligence, bias detection, and knowledge graphs. Moreover, an additional aim is to improve intent recognition in order to separate misinformation, disinformation, and malinformation better from each other. This distinction serves the purpose to enhance the detection of specific types of information disorders.

Keywords

Information disorder, detection, natural language processing, bias, knowledge graph, explainability

1. Introduction

Problem Statement and Relevance. The Global Risk Report 2024 [1] identifies *misinformation and disinformation* – on which this work places particular emphasis – as the number one global risk (ranked by severity over the short and long term) for the upcoming two years and on the fifth rank for the next ten years. Geopolitical events and developments are interconnected with information disorders (cf. [2]) and have led to societal polarization. Advancements in artificial intelligence (AI) (e.g., deep fakes) have only accelerate this.

Since 2008 online content sharing platforms (OCSPs) have been on the rise (cf. [3]). But this development and the change of information consumption have also caused problems. Types of information disorder (cf. [2]), as shown in *Figure 1*, which spread rapidly, have caused economic and societal harm. Examples for the latter are the COVID-19 pandemic or *infodemic*, like the World Health Organizations (WHO) denoted it alternatively (cf. [4]), the European Medicines Agency (EMA) hacking in 2020 and release of manipulated vaccine data (cf. [5]), as well as the 2021 United States Capitol attack (cf. [6]). However, for voters it is key to have access to authentic information or facts within a *networked society* (cf. [7]).

Proceedings of the Doctoral Consortium at ISWC 2024, co-located with the 23rd International Semantic Web Conference (ISWC 2024)

mark.nicolas.gruensteidl@wu.ac.at (M. N. Gruensteidl)
0009-0000-1427-4748 (M. N. Gruensteidl)

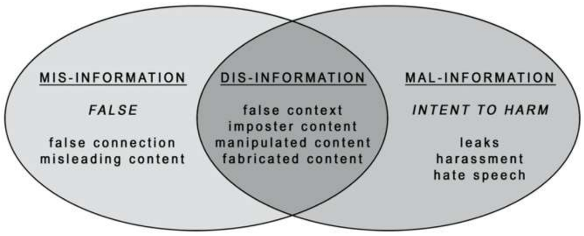

Figure 1: Information Disorder Types (cf. [2]).

That way a *monitorial citizen* (cf. [8]) can come to a *rational choice* in the sense that an individual follows his or her rational interest when voting (cf. [9]). However, information disorders make this task more difficult.

Policies and legal regulations have addressed different media content (e.g., text, images, etc.), OCSPs (e.g., data access), and flows of information (e.g., disinformation task forces) while at the same time protecting human rights and fundamental rights (cf. [7, 10, 11]). However, transparency and trust are open issues (cf. [12]), especially when it comes to solution approaches that should fulfill legal requirements as well as societal expectations (cf. [13]). While critical thinking and media literacy are indeed crucial skills, in order to establish societal resilience against information disorders (cf. [14]), information technology can assist democratic actors such as citizens with information disorder detectors (cf. [15]) and provide meaningful and transparent further context (cf. [16]) to voters for example. Additionally, there exists a lack of (training) data (cf. [17]). The consequence of this are possibly biased natural language processing (NLP) machine learning (ML) and deep learning (DL) models (cf. [18]). Lastly, information disorders are complex regarding their phases, specific types, multiple classes, content variability, rapid dissemination, context sensitivity, origins, dynamic nature, and when they are generated by AI (cf. [19]).

Focus and Contributions. The first focus lies on improving information disorder detection. For that, different detection algorithms (e.g., deep learning, BERT; rule or lexicon-based; linear regression, Support Vector Machine; probabilistic, Naïve Bayes; proximity-based, K-nearest neighbors algorithm – see *Figure 3*) will be used to evaluate and compare their performance for the task of detecting information disorder types on five selected datasets (see *Table 1*) with respect to mis-, dis-, and malinformation (depicted in *Figure 1*). A second research aim is the analysis of potential bias within black box models applying explainable artificial intelligence (XAI) tools. Examples for the latter include LIME (Local interpretable model-agnostic explanations) as well as Anchors (cf. [38]). Further, to enrich the contextual understanding of choices a model makes, a knowledge graph (KG) can be applied. But also existing fairness toolkits for bias detection are relevant for the analysis (cf. [20]). Third, the goal is to investigate to what extent it is possible to detect intent to harm. When it comes to transparency, a KG can add useful context to an intent recognition (IR) model. The overarching goal is to increase trust in AI-based detection models as well as help research, societal, and business stakeholders to cope with information disorders better.

Paper Structure. The rest of this paper is organized as follows: in Section 2 related work is presented and discussed. Afterwards, the research objectives, open issues or challenges, main hypothesis, research questions, methodology, research plan, and data sources are given in Sections 3 and 4. Finally, Section 5 concludes and identifies future work.

2. Related Work

Already when the COVID-19 pandemic started, AI fact-checking models were available. But training datasets did not specifically cover COVID-19 misinformation which led to domain analysis issues and limited the application scope of these detection models (cf. [21]). Therefore, domain-specific datasets, that are designed to include the unique characteristics of such events, had to be created (cf. [21, 22]). Another problem is that false information tends to spread faster compared to facts. Typical NLP classification models for misinformation detection are based on BERT or modified versions of it and misinformation can be analyzed looking into textual elements like sentiment or veracity (cf. [22]). The overall aim is to tackle it by detecting or validating its content and analyze its dynamics or management (cf. [23]). Regarding the analysis of the distribution of false information and datasets that are used to train detectors, specific topics or demographics matter (cf. [24]).

Others like Mensio and Alani [25] analyzed who interacts with misinformation. Interestingly, readability is better for false information and the more complex it is the more people believe in it. In this regard, characteristics like sentence and word length are relevant (cf. [26]). Bruel and Alani [27] crafted a reporting tool for dynamic spread and fact-checks. Detection of false claims is a multidisciplinary challenge and especially DL as well as XAI have been researched in computer science. It should be taken into account that combating information disorders is a task that can be tackled via human-centric artificial intelligence (AI) approaches good (cf. [28]). When it comes to large language models (LLMs), Leite et. al. [29] emphasized the time-consuming annotation process and role of credibility signals.

Although there have been significant advancements in terms of NLP-based text analysis tools, reputation evaluation, network data analyses, and imaged-based detection techniques, there are a number of open challenges with respect to the *concept drift* and *dynamic streaming nature* of (false) information but also deepfakes represent an arising problem. Explainable lifelong learning concepts could be a remedy (cf. [30]).

From a symbolic AI point of view, semantic web (SW) concepts like ontologies (e.g., OWL) (cf. [31]) and KGs have already been used to detect fake news. Tchechmedjiev et. al. [16] have shown an extensible KG consisting of fact-checked claims that serves to tackle the lack of corpora containing structured (meta) data. Pan et. al. [32] also used KGs for content based fake news detection and showed promising results for detecting false content. Moreover, they found that even incomplete or imprecise KGs are helpful.

Kaliyar et. al. [33] have shown in their experiments that BERT-based fake news detection can reach a high accuracy. Combining such DL models with SW technologies can be done in various ways. Denaux et al. [34] introduced a data model and distributed agents' architecture for composable credibility reviews in context of explainable misinformation detection. Furthermore, Lovera et. al. [35] took a different approach and worked on another hybrid solution. They analyzed the sentiment of short texts by using KGs and DL technologies to classify their sentiment. Traceability but also interpretability via XAI led to results that are more transparent. Such an approach compensates for the weaknesses of a black box model in this regard. Their model outperformed character n-gram solutions.

In particular, adapting original data sources with additional information from semantic resources is prominent for contextual enrichment. Some KGs are commonly used in that regard (e.g., ConceptNet or DBpedia) (cf. [18]). KGs can help to identify relevant events for information disorder detection as certain events are often linked to fake news. Opdahl and Tessem [36] looked into ontologies to find journalistic angles. KGs can also support detecting online toxicity by mitigating possible bias, subjective views, or lacking domain knowledge of annotators (cf. [37]). Keeping in mind that black box models should be explainable, Szczepański et. al. [38] proposed an extension which is based on LIME and Anchors that enriches a model's classification with additional reasoning. XAI techniques like this and KGs for facilitating context-based detection of information disorders like *Fane-KG* (cf. [39]) are on the rise. *Fane-KG* represents a KG that is designed to be used specifically for context-based fake news detection. Hani et. al. [39] provide a human as well as machine processable data repository built on semantic standards (OWL and RDF).

Generating a KG for news articles, which can be then used for detecting or checking false information, is challenging. Generally speaking, KG construction consists of three phases: the acquisition, refinement, and evolution of knowledge (cf. [40]). Such created KGs can then be used to improve the detection of information disorders. Mayank et. al. [41] worked on a combination of a NLP and tensor decomposition model. They encoded content from news and embedded KG entities. The latter and their variety could improve their detector. There is a research trend to combine ML and SW technologies like KGs identifiable (cf. [42]).

Lastly, regarding intent recognition (IR), there are already pretrained BERT-based models that achieve promising and precise results (cf. [43]). In addition, domain-specific KGs can improve the intent classification task (cf. [44]). Zhou et. al. [45] assessed the intent of shared fake news and proposed an influence graph as well as modeled the intent of an individual when spreading fake news. Furthermore, a centralized KG with checked articles could help with queries and add contextual knowledge for fact-checking tasks (cf. [46]).

3. Hypothesis and Research Questions

Starting with the challenges, as outlined in Section 2, AI techniques (i) used to identify information disorders are not transparent enough and (ii) lack in cross-domain applicability. In addition, (iii) AI causes issues if producing deep fakes (cf. [13]). Researchers classify *fake news* mostly based on intent (e.g., to harm) and content (e.g., text or images). Also, (iv) it is a challenge to label *fake news* since there are many terms (cf. [2]). According to Farhangian et. al. [47], it may be useful for information disorder detection research purposes to look into multiple feature representations, perspectives, dynamic ensemble models, and a cross-dataset evaluation (tackling concept drift). DL and ML can identify domain-specific information disorders but there exists a lack of trust into black box models (cf. [13]). Furthermore, (v) missing data for detection models is a major issue (cf. [17]) and (vi) data validity (e.g., incomplete data) or how humans use AI systems are problems that can increase bias in algorithms (cf. [18]). Combining symbolic and sub-symbolic AI has the potential to fill those weaknesses of current AI systems. Common approaches to mitigate bias are altering the data AI systems are trained on, changing how a model learns, or using a holdout set (not part of training) to adapt the outcome. However, there is still not enough research on identifying bias (cf. [18]). Additionally, combining a NLP DL approach with a KG, using structured and unstructured data, can outperform non-hybrid models (cf. [48]).

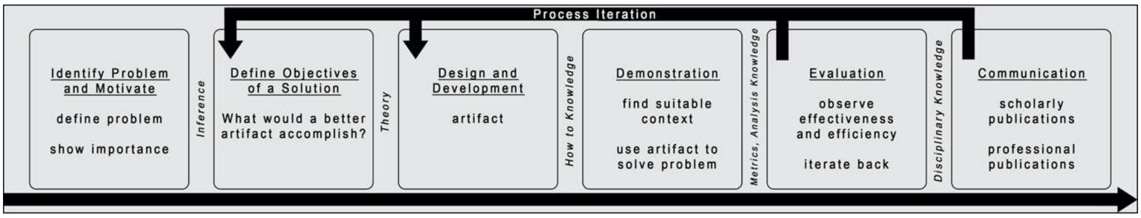

Figure 2: DSRM Cycle; adapted from Peffers et al., 2007 (cf. [49, 50]).

Regarding XAI, Yuan et. al. [28] concluded that explainable detection of false information from a user perspective is key and highlighted human-machine communication. As a last obstacle for detecting and analyzing information disorders, there remains (vii) the open issue of classifying it better into mis-, dis-, and malinformation. For this task, automatically created domain-specific KGs can assist during the intent classification process (cf. [44]) and add for example useful contextual information (cf. [46]). For instance, provenance or temporal characteristics of information disorders could be considered in this context.

Therefore, the main hypothesis for this research proposal is summarized below:

A more transparent information disorder detection model may be made possible by: (i) a domain-specific cross-dataset analysis for information disorder detection using ML and DL; (ii) the enrichment of ML and DL with knowledge graph-based meaning that facilitates explainability and bias detection; and (iii) intent recognition approaches that can distinguish between mis-, dis-, and malinformation.

The following research questions (RQs) are derived from this hypothesis:

RQ1) Which machine and deep learning algorithms are most effective when it comes to information disorder detection?

RQ2) To what extent can machine and deep learning explainability and bias detection be facilitated via knowledge graph-based enrichment?

RQ3) How can hybrid AI-based approaches be used to better distinguish between the three information disorder types (misinformation, disinformation, and malformation)?

4. Research Plan and Methodology

This research project is planned according to the design science research methodology (DSRM), which is an information systems specific research method that aims to develop innovative artifacts (cf. [51]) and consists of six phases as shown in *Figure 2* (cf. [49, 50]).

The **problem and motivation** is to increase transparency and trust in AI-based detection models in order to help stakeholders to better cope with information disorders. The primary goal is to improve the detection of information disorders, explainability of models, bias detection as well as mitigation, and intent recognition.

In terms of **solution objectives**, the plan is to combine ML or DL algorithms with XAI, bias detection, and KGs, in order to understand and mitigate potential bias in information disorder detection models better. However, also for intent recognition, which is needed to distinguish the three information disorder types (misinformation, disinformation, and malinformation) effectively, transparency is key. Addressing the lack of training datasets (cf. [28]), one of the envisaged outcomes and innovative artifacts (cf. [51]) is a new dataset incorporating the three information disorder types with binary as well as non-binary labels.

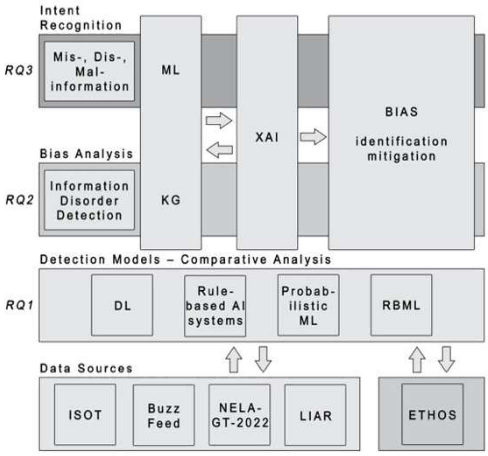

Figure 3: Research Plan.

Content	Dataset	Source
Fake News	ISOT	[52] [53]
	BuzzFeed	[54]
	LIAR	[55]
Misinformation	NELA-GT-2022	[56]
Hate Speech	ETHOS	[57]

Table 1: Information Disorder Datasets.

Design and Development aims to create one or more prototypes as artifacts that address the issues highlighted under the solution objectives heading (cf. [50]). *Figure 3* shows the research plan. Possible artifacts include (i) statistical insights (accuracy, recall, precision, F1 score), (ii) performance results, (iii) models, (iv) datasets, or (v) KGs. In addition, *Table 1* shows five selected datasets which cover different types of information disorders.

During **demonstration**, the artifact is used to address the problem or instances derived from a concrete use case scenario. This can be a case study or experimental setting (cf. [50]). For RQ1, this includes the application of detection model prototypes. Regarding RQ2, KGs and statistical insights from XAI are applied to analyze bias. Third, for RQ3, a new dataset containing mis-, dis-, and malinformation as well as a model are part of the demonstration.

In order to **evaluate** the artifact, we need to analyze how well it provides a solution to the issue(s) it aims to address. Using selected quantitative metrics (e.g., accuracy, precision, recall, F1 score, etc.), the objectives and goals can be compared as well as analyzed (cf. [50]).

Finally, **communication** ensures that the problem and its relevance, artifacts and their utility, consequent design, and solutions as well as their effectiveness are presented to our stakeholders. The plan is to publish in highly ranked conferences and journals (cf. [50]).

5. Conclusions and Future Work

This proposal aims to improve the detection of information disorders, bias detection and mitigation, intent recognition, and stressed the need for more transparency and explainability concerning AI detection models. We started by investigating the status quo, discussed research questions, described how to approach open issues, and showed the research plan as well as methodology guiding it. A limitation of this work is that it does not address multi-modal detection. Future work includes tackling the research questions with these key objectives: (i) a cross-dataset and detection model comparison; (ii) a bias analysis deploying XAI and KGs; and (iii) the separation of information disorders by intent to harm.

Acknowledgements

This research is conducted under the supervision of Prof. Sabrina Kirrane.

References

[1] World Economic Forum, The Global Risks Report 2024, World Economic Forum, 2024. URL: https://www.weforum.org/publications/global-risks-report-2024/.
[2] C. Wardle, H. Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe, 2017.
[3] E. Ortiz-Ospina, The rise of social media, Our World in Data, 2019. URL: https://ourworldindata.org/rise-of-social-media.
[4] United Nations, UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis, 2020. URL: https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19.
[5] L. Cerulus, EU medicines agency says hackers manipulated leaked coronavirus vaccine data, Politico, 2021. URL: https://www.politico.eu/article/european-medicines-agency-ema-cyberattack-coronavirus-vaccine-data/.
[6] L.W. Green, J.E. Fielding, R.C. Brownson, More on Fake News, Disinformation, and Countering These with Science, Annu Rev Public Health 42 (2021) vi. https://doi.org/10.1146/annurev-pu-42-012821-100001.
[7] P. Iosifidis, N. Nicoli, Digital Democracy, Social Media and Disinformation, Milton, 2020. https://doi.org/10.4324/9780429318481.
[8] A. Chadwick, Web 2.0: New Challenges for the Study of E-Democracy in an Era of Informational Exuberance, I/S: A Journal of Law and Policy for the Information Society 5 (2009) 9–41.
[9] R. Antunes, Theoretical models of voting behaviour, Exedra 4 (2010) 145–170.
[10] C. COLOMINA, H. SÁNCHEZ MARGALEF, R. YOUNGS, The impact of disinformation on democratic processes and human rights in the world, European Parliament - Policy Department, 2021. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf.
[11] D. Vese, Governing Fake News: The Regulation of Social Media and the Right to Freedom of Expression in the Era of Emergency, European Journal of Risk Regulation (2021) 41. https://doi.org/10.1017/err.2021.48.
[12] Fabrizio Di Mascio, Michele Barbieri, Alessandro Natalini, Donatella Selva, Covid-19 and the Information Crisis of Liberal Democracies: Insights from Anti-Disinformation Action in Italy and EU, Partecipazione e Conflitto 14 (2021) 240. https://doi.org/10.1285/i20356609v14i1p221.
[13] E. Aimeur, S. Amri, G. Brassard, Fake news, disinformation and misinformation in social media: a review, Social Network Analysis and Mining 13 (2023). https://doi.org/10.1007/s13278-023-01028-5.
[14] S.M. Jones-Jang, T. Mortensen, J. Liu, Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't, AM BEHAV SCI 65 (2021) 388. https://doi.org/10.1177/0002764219869406.
[15] T. Cassauwers, Can artificial intelligence help end fake news?, Horizon Magazine, 2019. URL: https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news.
[16] A. Tchechmedjiev, P. Fafalios, K. Boland, M. Gasquet, M. Zloch, B. Zapilko, S. Dietze, K. Todorov, ClaimsKG: A Knowledge Graph of Fact-Checked Claims, in: C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. Cruz, A. Hogan, J. Song, M. Lefrançois, F. Gandon (Eds.), The Semantic Web – ISWC 2019, Springer International Publishing, Cham, 2019: pp. 309–324.

[17] K. Hao, Even the best AI for spotting fake news is still terrible, MIT Technology Review, 2018. URL: https://www.technologyreview.com/2018/10/03/139926/even-the-best-ai-for-spotting-fake-news-is-still-terrible/.
[18] P. Reyero, E. Daga, H. Alani, M. Fernández, A Systematic Survey of Semantic Web Technologies for Bias in Artificial Intelligence Solutions, in: 2021. https://api.semanticscholar.org/CorpusID:237102597.
[19] S. Tufchi, A. Yadav, T. Ahmed, A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities, International Journal of Multimedia Information Retrieval 12 (2023). https://doi.org/10.1007/s13735-023-00296-3.
[20] B. Johnson, J. Bartola, R. Angell, S. Witty, S. Giguere, Y. Brun, Fairkit, fairkit, on the wall, who's the fairest of them all? Supporting fairness-related decision-making, EURO Journal on Decision Processes 11 (2023) 100031. https://doi.org/10.1016/j.ejdp.2023.100031.
[21] Y. Jiang, X. Song, C. Scarton, A. Aker, K. Bontcheva, Categorising Fine-to-Coarse Grained Misinformation: An Empirical Study of COVID-19 Infodemic, CoRR abs/2106.11702 (2021). https://arxiv.org/abs/2106.11702.
[22] Y. Peskine, R. Troncy, P. Papotti, Analyzing COVID-Related Social Discourse on Twitter using Emotion, Sentiment, Political Bias, Stance, Veracity and Conspiracy Theories, in: Companion Proceedings of the ACM Web Conference 2023, Association for Computing Machinery, New York, NY, USA, 2023: pp. 688–693. https://doi.org/10.1145/3543873.3587622.
[23] M. Fernandez, H. Alani, Online Misinformation: Challenges and Future Directions, in: Companion Proceedings of the The Web Conference 2018, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018: pp. 595–602. https://doi.org/10.1145/3184558.3188730.
[24] G. Burel, T. Farrell, H. Alani, Demographics and topics impact on the co-spread of COVID-19 misinformation and fact-checks on Twitter, Information Processing & Management 58 (2021) 102732. https://doi.org/10.1016/j.ipm.2021.102732.
[25] M. Mensio, H. Alani, MisinfoMe: Who is Interacting with Misinformation?, in: M.C. Suárez-Figueroa, G. Cheng, A.L. Gentile, C. Guéret, C.M. Keet, A. Bernstein (Eds.), Proceedings of the ISWC 2019 Satellite Tracks (Posters & Demonstrations, Industry, and Outrageous Ideas) Co-Located with 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, October 26-30, 2019, CEUR-WS.org, 2019: pp. 217–220. https://ceur-ws.org/Vol-2456/paper57.pdf.
[26] M. ali Tavakoli, H. Alani, G. Burel, On the Readability of Misinformation in Comparison to the Truth, in: Text2Story@ECIR, 2023. https://api.semanticscholar.org/CorpusID:258334770.
[27] G. Burel, H. Alani, The Fact-Checking Observatory: Reporting the Co-Spread of Misinformation and Fact-checks on Social Media, in: Proceedings of the 34th ACM Conference on Hypertext and Social Media, Association for Computing Machinery, New York, NY, USA, 2023. https://doi.org/10.1145/3603163.3609042.
[28] L. Yuan, H. Jiang, H. Shen, L. Shi, N. Cheng, Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice, Systems 11 (2023) 458. https://doi.org/10.3390/systems11090458.
[29] J.A. Leite, O. Razuvayevskaya, K. Bontcheva, C. Scarton, Detecting misinformation with llm-predicted credibility signals and weak supervision, arXiv Preprint arXiv:2309.07601 (2023).
[30] M. Choraś, K. Demestichas, A. Giełczyk, Á. Herrero, P. Ksieniewicz, K. Remoundou, D. Urda, M. Woźniak, Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study, Applied Soft Computing 101 (2021) 107050. https://doi.org/10.1016/j.asoc.2020.107050.

[31] M. Lahby, A.-S. Khan Pathan, Y. Maleh, W.M.S. Yafooz, Combating Fake News with Computational Intelligence Techniques, 1st ed., Springer Cham, 2021.
[32] J.Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, J. Liu, Content Based Fake News Detection Using Knowledge Graphs, in: International Workshop on the Semantic Web, 2018. https://api.semanticscholar.org/CorpusID:52900831.
[33] R.K. Kaliyar, A. Goswami, P. Narang, FakeBERT: Fake news detection in social media with a BERT-based deep learning approach, Multimedia Tools and Applications 80 (2021) 11765–11788. https://doi.org/10.1007/s11042-020-10183-2.
[34] R. Denaux, M. Mensio, J.M. Gomez-Perez, H. Alani, Weaving a Semantic Web of Credibility Reviews for Explainable Misinformation Detection (Extended Abstract), in: Z.-H. Zhou (Ed.), Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, International Joint Conferences on Artificial Intelligence Organization, 2021: pp. 4760–4764. https://doi.org/10.24963/ijcai.2021/646.
[35] F.A. Lovera, Y.C. Cardinale, M.N. Homsi, Sentiment Analysis in Twitter Based on Knowledge Graph and Deep Learning Classification, Electronics 10 (2021). https://doi.org/10.3390/electronics10222739.
[36] A.L. Opdahl, B. Tessem, Ontologies for finding journalistic angles, Software and Systems Modeling 20 (2021) 71–87. https://doi.org/10.1007/s10270-020-00801-w.
[37] P. Reyero Lobo, E. Daga, H. Alani, Supporting Online Toxicity Detection with Knowledge Graphs, Proceedings of the International AAAI Conference on Web and Social Media 16 (2022) 1414–1418. https://doi.org/10.1609/icwsm.v16i1.19398.
[38] M. Szczepański, M. Pawlicki, R. Kozik, M. Choraś, New explainability method for BERT-based model in fake news detection, Scientific Reports 11 (2021). https://doi.org/10.1038/s41598-021-03100-6.
[39] A. B. Hani, O. Adedugbe, F. Al-Obeidat, E. Benkhelifa, M. Majdalawieh, Fane-KG: A Semantic Knowledge Graph for Context-Based Fake News Detection on Social Media, in: 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), 2020: pp. 1–6. https://doi.org/10.1109/SNAMS52053.2020.9336542.
[40] L. Zhong, J. Wu, Q. Li, H. Peng, X. Wu, A Comprehensive Survey on Automatic Knowledge Graph Construction, ACM Comput. Surv. 56 (2023). https://doi.org/10.1145/3618295.
[41] M. Mayank, S. Sharma, R. Sharma, DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection, 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2021) 47–51.
[42] A. Breit, L. Waltersdorfer, F.J. Ekaputra, M. Sabou, A. Ekelhart, A. Iana, H. Paulheim, J. Portisch, A. Revenko, A.T. Teije, F. Van Harmelen, Combining Machine Learning and Semantic Web: A Systematic Mapping Study, ACM Comput. Surv. 55 (2023). https://doi.org/10.1145/3586163.
[43] V. Khan, T.A. Meenai, Pretrained Natural Language Processing Model for Intent Recognition (BERT-IR), Human-Centric Intelligent Systems 1 (2021) 66–74. https://doi.org/10.2991/hcis.k.211109.001.
[44] M. Arcan, S. Manjunath, C. Robin, G. Verma, D. Pillai, S. Sarkar, S. Dutta, H. Assem, J.P. McCrae, P. Buitelaar, Intent Classification by the Use of Automatically Generated Knowledge Graphs, Information 14 (2023). https://doi.org/10.3390/info14050288.
[45] X. Zhou, K. Shu, V.V. Phoha, H. Liu, R. Zafarani, "This is Fake! Shared it by Mistake":Assessing the Intent of Fake News Spreaders, in: Proceedings of the ACM Web Conference 2022, Association for Computing Machinery, New York, NY, USA, 2022: pp. 3685–3694. https://doi.org/10.1145/3485447.3512264.
[46] G.K. Shahi, FakeKG: a knowledge graph of fake claims for improving automated fact-checking (student abstract), in: Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial

Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI Press, 2023. https://doi.org/10.1609/aaai.v37i13.27020.
[47] F. Farhangian, R. Cruz, G. Cavalcanti, Fake news detection: Taxonomy and comparative study, Information Fusion 103 (2023) 102140. https://doi.org/10.1016/j.inffus.2023.102140.
[48] A.M.P. Braşoveanu, R. Andonie, Semantic Fake News Detection: A Machine Learning Perspective, in: I. Rojas, G. Joya, A. Catala (Eds.), Advances in Computational Intelligence, Springer International Publishing, Cham, 2019: pp. 656–667.
[49] J. Gregório, L. Reis, M. Peyroteo, M. Maia, M. Mira da Silva, L.V. Lapão, The role of Design Science Research Methodology in developing pharmacy eHealth services, Research in Social and Administrative Pharmacy 17 (2021) 2089–2096. https://doi.org/10.1016/j.sapharm.2021.05.016.
[50] K. Peffers, T. Tuunanen, M. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of Management Information Systems 24 (2007) 45–77.
[51] A.R. Hevner, S.T. March, J. Park, S. Ram, Design Science in Information Systems Research, MIS Quarterly 28 (2004) 75–105. https://doi.org/10.2307/25148625.
[52] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classification, Security and Privacy 1 (2017) e9. https://doi.org/10.1002/spy2.9.
[53] H. Ahmed, I. Traore, S. Saad, Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques, 2017. https://doi.org/10.1007/978-3-319-69155-8_9.
[54] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A Stylometric Inquiry into Hyperpartisan and Fake News, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018: pp. 231–240. https://doi.org/10.18653/v1/P18-1022.
[55] W.Y. Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, in: R. Barzilay, M.-Y. Kan (Eds.), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Vancouver, Canada, 2017: pp. 422–426. https://doi.org/10.18653/v1/P17-2067.
[56] M.G. Gruppi, B.D. Horne, S. Adali, NELA-GT-2022: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles, in: 2022. https://api.semanticscholar.org/CorpusID:260439925.
[57] I. Mollas, Z. Chrysopoulou, S. Karlos, G. Tsoumakas, ETHOS: a multi-label hate speech detection dataset, Complex & Intelligent Systems 8 (2022) 4663–4678. https://doi.org/10.1007/s40747-021-00608-2.