Al-Noor Journal for Digital Media Studies
October, 2025; v.2, n.1, a.4 (39-47)
ISSN: 3007-9853, DOI: https://doi.org/10.69513/jnfdms.v2.i1.a4

Research Article

Using Artificial Intelligence to Detect and Mitigate Fake News

Alexandra M. Kasimova[a*]

2241133074@pfur.ru

ORCID:0009-0004-4193-9935

Ekaterina S. Kozlovskaya[b]

kozlovskaya_es@pfur.ru

ORCID: 0000-0002-6308-725X

[ab]RUDN University, Miklukho-Maklaya str., 6, Moscow, 117198, Russia

Received: 1/08/2025 Accepted: 10/09/2025 Published:11/10/2025

Abstract:

The article examines the role of Artificial Intelligence (AI) in combating fake news, exploring its methods, successful applications, and key benefits. AI technologies, such as machine learning and natural language processing, enable the rapid detection of false information by analyzing patterns and inconsistencies in large datasets. Examples of successful AI implementation include social media platforms like Facebook and Twitter, which use algorithms to identify and flag misleading content. The advantages of AI lie in its speed, scalability, and ability to process information in real time, making it a critical tool in the fight against disinformation.

However, the article also highlights the technical, ethical, and social risks associated with AI. Technical challenges include algorithmic bias and the difficulty of detecting sophisticated deepfakes. Ethical concerns revolve around privacy, censorship, and the potential misuse of AI for propaganda. Social risks involve the erosion of trust in media and the exacerbation of polarization. To address these challenges, the article emphasizes the need for an interdisciplinary approach, combining expertise in technology, ethics, law, and social sciences to develop robust information verification systems.

The study concludes that while AI is a powerful tool for countering fake news, it also poses significant risks if not properly regulated. To maximize its benefits, strict control, transparency, and adherence to ethical standards are essential.

* Corresponding author
E-mail address: 1132243074@pfur.ru

Al-Noor Journal for Digital Media Studies
October, 2025; v.2, n.1, a.4 (39-47)
ISSN: 3007-9853, DOI: https://doi.org/10.69513/jnfdms.v2.i1.a4

Recommendations include fostering collaboration between technologists, policymakers, and ethicists to create frameworks that ensure responsible AI use. Additionally, promoting media literacy among the public can help mitigate the impact of disinformation. The future of AI in this field depends on balancing innovation with accountability, ensuring that AI contributes positively to the information landscape.
In summary, AI has the potential to revolutionize the fight against fake news, but its application must be carefully managed to avoid unintended consequences. By addressing technical limitations, ethical concerns, and social risks, and by adopting an interdisciplinary approach, society can harness the power of AI to create a more informed and resilient information ecosystem.

Keywords: artificial intelligence, fake news, information, verification, deepfake, media literacy, machine learning algorithms, transparency, ethics, technology regulation

Introduction

The competition of transnational corporations in modern conditions of globalization requires a deep analysis of the factors that ensure their effectiveness. The studied topic considers key aspects reflecting the main criteria for evaluating the activities of TNCs – their innovative potential, financial stability, adaptability to changing conditions, as well as comprehensive strategies implementing modern management and marketing methods. The systematic combination of these parameters creates a solid foundation for the corporation's long-term positions in the international arena (Bebich, D., 2018: 450).

The undeniable value in this process is played by innovations that allow TNCs to remain on the crest of success. The introduction of advanced technologies and constant monitoring of the needs of the audience lead to the creation of unique products and services, providing the company with an economic advantage. An example is Samsung and Apple, which present high-tech products, focusing on modern market requirements.

Economic and structural factors, including financial literacy and logistics, also significantly influence the development of TNC competitiveness. Toyota and Amazon demonstrate how the coordinated work of financial services and set processes determines their success in the international arena, confirming the relevance of implementing effective management models to achieve commercial goals and sustainable development.

This study is devoted to the analysis of aspects of the use of intelligent systems in the selection and verification of information, with an emphasis on

the risks and limitations that may arise in the process. It is impossible to ignore the socio-cultural, technical and legal factors that directly affect the effectiveness of the introduction of modern technologies.

One of the most disturbing manifestations is the growing popularity of synthetic media content, in particular, deepfake technologies that allow to create fake videos. Analysts' forecasts show that in 2022, over eighty thousand such materials were registered, and the annual increase reaches sixty percent. This trend raises serious concerns, undermining the rating reputation of the media space and leaving regulators with difficult tasks for productive control over content (Bortnik, A.D., 2022:86).

Based on the above, it should be emphasized that artificial intelligence should not be considered a panacea for disinformation. Only if agreed solutions are reached to overcome existing challenges, the use of AI can really take the role of an essential element in the fight against fake materials and manipulation of public opinion.

Success in this area is possible only in the light of the integration of advanced technologies, adequate legislative regulation and increased critical perception of information among users. Taking into account these factors, the study will be based on a detailed study.

The potential of artificial intelligence in the fight against fake news

In recent years, artificial intelligence (AI) has become an important tool for recognizing fake news and disinformation. The main advantage of modern solutions is the ability to automatically analyze large amounts of information at high speed.

Modern machine learning methods, including various algorithmic approaches, are able to recognize certain signs of fake news. Thus, the work of a group of scientists led by Jahanzaeb Anwar in 2022 demonstrated that the use of Random Forest and Support Vector Machines algorithms makes it possible to achieve 85% accuracy in recognizing disinformation in texts posted on social networks. Modern algorithms are able to analyze the linguistic and structural characteristics of a text, including the analysis of the use of emotionally colored words, complex syntactic constructions and links to unreliable resources (Davydov, S.G., 2023: 141).

Neural network approaches in AI show their effectiveness in processing visual content. The generation of artificial media, such as deepfake videos,

requires new methods to identify deceptive content. As part of the Deepfake Detection Challenge initiative, organized by Facebook and Microsoft in 2020, the ability of neural networks based on the CNN architecture to detect 65% of fakes in video materials, revealing manipulation of public opinion, such as the incident with Nancy Pelosi in 2019, which received huge coverage, was demonstrated.

Neuro-linguistic models, including GPT and BERT, accelerate the processes of contextual analysis and confirmation of information. Having access to innovative tools significantly strengthens the protective mechanisms against disinformation and contributes to improving the media literacy of the population.

Deep machine learning algorithms provide the ability to process vast amounts of textual information by comparing and verifying data with existing databases of reliable information. In 2021, the platform FactCheck.org I applied the BERT-based methodology in my systems to verify the actual information, which made it possible to reduce time costs by 30% when verifying the content of various requests.

Among the resources that have implemented similar solutions, Snopes and Full Fact stand out, using AI for automated analysis of political statements, news headlines and even Internet memes, which improves the reliability of information in actively discussed political events.

Social media such as Facebook, Twitter and YouTube are actively incorporating AI capabilities to counter the spread of false information. As of 2023, the Facebook platform reported that the AI tools used for instant content analysis continue to show impressive results: more than 8 billion falsified messages were deleted automatically, of which 98% of the cases were carried out before users complained about fakes (Sukhodolov, A.P., 2017: 652).

The regular use of tools such as artificial intelligence (AI) creates barriers to the rapid spread of disinformation, which significantly weakens its potential impact on public opinion. The primary advantage of automated systems is the ability to process countless amounts of information, which is especially important in conditions of constant content flow.

As Google confirms, its trained algorithms can handle the analysis of up to 20 million new articles on each active working day, selecting the most critical ones for further detailed analysis. The undeniable advantage of AI in this area is the absence of bias inherent in human judgments. Excluding the influence of

Al-Noor Journal for Digital Media Studies
October, 2025; v.2, n.1, a.4 (39-47)
ISSN: 3007-9853, DOI: https://doi.org/10.69513/jnfdms.v2.i1.a4

emotional and political factors on the results makes the information verification process less subjective. At the same time, the resulting effectiveness of using such tools depends entirely on the quality of the initial data of their training, which implies the need for regular evaluation and updating of the algorithms used.

Risks and limitations of using AI

Despite the promising prospects of artificial intelligence (AI) in the field of neutralizing disinformation, the introduction of intelligent technologies is associated with a number of dangers and difficulties. The risk of misinterpretation of information content remains among the primary difficulties. The effectiveness of machine learning tools is directly related to the qualitative criteria and the scale of the source data. Thus, the results of Monica Riolli's work (2021) demonstrated that every fifth exhaustive analysis of materials was observed with a distortion of conclusions due to the uneven distribution of information. Parallels between objective factors and their subjective interpretations create multidisciplinary difficulties in recognizing relevant facts (Tretyakov, A.O., 2018: 567).

Anomalies in the definition of news contribute to the formation of user distrust of the proposed platforms and automatic solutions. Strict restrictions on data processing represent a significant obstacle. Improving legal control mechanisms similar to the European GDPR standards requires research laboratories and private companies to find alternative methods of working with information flows. In particular, in 2020, the Google AI group faced a shortage of local information resources when designing software solutions for verifying news content. This circumstance caused a drop in the functionality of the algorithmic software to 23%, comparison with similar initiatives on the American continent gave more optimistic results.

The implementation of ethical standards in the field of artificial intelligence requires special attention and regulation. It is known that resources designed to monitor fakes are used by hackers to create them. The generation of fake video materials, as well as crypto images created on the basis of generative neural networks, indicates a high level of threat to society in case of unauthorized use of AI (Gorokhov, A.V., 2022: 5).

According to the 2019 Deeptrace report, the number of deepfake videos increased by 84% compared to the previous year, and 96% of them used manipulative techniques. As an example, the target audience of fake videos

featuring world leaders such as Barack Obama or Donald Trump is focused on more than a million views before their forgery is confirmed.

Aspects of personal data protection call into question the ethics of using machine learning systems. The use of AI requires the accumulation of large amounts of information, including user data, which can lead to appeals to the jurisdiction without the knowledge of the owners of personal information.

The scandal surrounding Clearview AI in 2021, when the company used over three billion images without user consent, illustrates the seriousness of the problem. Public interest has increased significantly about privacy in artificial intelligence since this publication.

Social issues and legal challenges have attracted the attention of analysts to the topic. Trust in algorithmic solutions remains at the forefront of scientific discussions (Sukhodolov, A.P., 2019:101).

A 2022 Pew Research Center study found that almost half of the respondents do not trust algorithmic solutions. The problem is compounded in developing countries with low digital literacy, where the likelihood of spontaneous algorithm errors can be perceived as intricate actions.

The development of legal norms and standards in the field of AI is a difficult task. In 2021, the European jurisdiction launched the draft "Act on Artificial Intelligence", focused on more careful control over those areas where the risk of abuse is high. However, different international regulators cannot agree on common standards, which creates legal inconsistencies.

The Chinese legislative initiative is aimed at creating content monitoring mechanisms, but the approach significantly contrasts with Western realities, which definitely complicates multilateral partnerships in this area.

Thus, despite the enormous potential of artificial intelligence, its use requires a careful balance between technological capabilities, ethical principles and legal norms in order to avoid new threats and maintain public trust.

Recommendations and prospects

Improving artificial intelligence to combat disinformation requires a comprehensive analysis of technology, setting ethical standards and supporting progress in this area.

The development of algorithms that are open to study provides a basis for increasing trust between AI systems and users. An effective example is the OpenAI initiative, which provides an opportunity to visualize the work of language models. A detailed interpretation of the internal mechanisms provides

Al-Noor Journal for Digital Media Studies
October, 2025; v.2, n.1, a.4 (39-47)
ISSN: 3007-9853, DOI: https://doi.org/10.69513/jnfdms.v2.i1.a4

an understanding of the logic of decision-making algorithms, which, in turn, reduces the likelihood of errors.

Along with this, the synergy of AI and human expertise contributes to obtaining reliable information. The Partnership on AI study confirmed that the combination of analysis by algorithms and additional verification by specialists leads to increased accuracy of the result. The Full Fact platform has demonstrated the effectiveness of AI pre-calculations, after which the researchers make the final assessment. This minimizes the possible risks associated with automated systems (Ilyakhina, A.A., 2024: 585).

Emerging ethical principles and norms at the international level are necessary to develop effective recommendations on the use of AI. In 2021, the EU initiated the creation of a legislative framework with the help of the "Artificial Intelligence Act", which takes into account interaction and the fight against fakes in the media. UNESCO also emphasizes the need for commitment to human rights and openness in the implementation of AI. The sustainable development of legislative initiatives will help to avoid legal inconsistencies and improve the controllability of the practices of using analysis algorithms.

Increasing the level of media literacy among citizens is becoming an increasingly important task. It is necessary to integrate modules in educational standards focused on the development of analytical thinking and understanding of the principles of functioning of AI technologies.

For example, an initiative launched in Finland in 2023, called "AI Literature for All", reached more than 60% of the adult population, offered courses on detecting disinformation and the mechanics of AI. These measures form adaptive structures of society to face manipulation of public opinion.

Modern advances in AI are opening up new horizons in countering fake information flows. Integration systems that allow content analysis on leading platforms are already being developed. Using the example of Twitter, which introduced GPT–4-based algorithmic methods in 2024 to alert users about unfair content, there was a 38% decrease in interaction with such records.

In the future, such technology will be able to work in real time, providing instant analysis of news publications and prompt warning of potential misinformation.

Hybrid models synthesizing algorithmic and analytical approaches reliably show advantages. Data Research & The Society, conducted in 2022,

confirmed that combined methods of analysis form an increased level of trust among consumers of information (Golovatskaya, O.E., 2019: 60).

In the context of innovative progress in the field of AI, it is necessary to rely on the prism of ethical responsibility, interstate interaction and human involvement in decision-making processes, minimizing negative aspects and enhancing social effects.

Conclusion

Modern artificial intelligence is a powerful tool in the fight against the spread of false information, contributing to detailed analysis and fact-checking in the face of increasing amounts of data in the digital space. Artificial intelligence is able to quickly process large amounts of information, detect specific signs of fake news and operate in real time, which significantly increases its importance in the current fight against disinformation. The implementation of the effective application of artificial intelligence in this field requires overcoming several challenges related to technological limitations, ethical issues and legal aspects. One of the main conclusions is the critical importance of minimizing the risks associated with the possibility of misinterpretation of information, bias in algorithms and insufficient accessibility to a variety of data. Incorrect algorithmic decisions can undermine audience confidence and deepen the problem of confusion when systems train on insufficiently prepared or biased samples. For the effective functioning of AI technologies, it is necessary to implement more open platforms that demonstrate the effectiveness of work and explain the decisions made. This point is especially relevant in the social sphere, where inaccurate actions can lead to serious consequences. The ethical aspects of developing technologies always require rigorous analysis, since methods developed to counteract disinformation can also be used in its production (Shirin, D.I., 2023: 106).

The urgency of adapting to the new realities of the digital age requires a more stringent framework for the development of technologies that recycle visual content. The rapid growth in the number of videos with fake content and algorithmically created materials requires careful monitoring at the level of international institutions that ensure standardization. Thus, the approaches proposed by the structure responsible for regulating AI in Europe serve as a good start, but the crucial problem of coordinating initiatives on a global scale remains.

The education system also needs to be rethought, because media literacy is becoming a key skill to resist manipulation. Real educational startups form the ability of citizens to critically evaluate information and share their findings, which in turn creates a conscious creative society.

References:

1. Bebich, D. (2018). New problems – old solutions? A critical look at the report of the European Commission's High-level Expert Group on fake news and online disinformation. *Bulletin of the Peoples' Friendship University of Russia. Series: Political Science*, *20* (3), 447-460.
2. Bortnik, A. D. (2022). How artificial intelligence will change the world of the media market. *Bulletin of the Magistracy*, *12-2* (135), 84-87.
3. Davydov, S. G. (2023). The use of artificial intelligence technologies in Russian media and journalism. *Bulletin of the Moscow University. Episode 10. Journalism*, *(5)*, 3-21. doi: 10.30547/vestnik.journ.5.2023.321
4. Golovatskaya, O. E. (2019). The meaning and origin of the term "fake news". *Communication*, *7* (2), 139-152.
5. Gorokhov, A.V. (2022). Artificial intelligence. *Skif. Student science issues*, *4* (68), 159-162.
6. Ilyakhina, A. A**.** (2024). Prospects for the use of artificial intelligence technologies in journalism. *Bulletin of Science*, *3*, 1 (70), 580-588.
7. Shirin, D. I. (2023). The impact of artificial intelligence on the modern world. *Science and Education*, *4* (4), 564–570.
8. Sukhodolov, A.P. (2017). The phenomenon of "Fake news" in the modern media space. *Eurasian Cooperation: Humanitarian Aspects*, *1*, 87-106.
9. Sukhodolov, A.P. (2019). Journalism with artificial intelligence. *Questionsoftheoryandpracticeofjournalism*, *8* (4), 647-667.
10. Tretyakov, A.O**.** (2018). A method for determining Russian-language fake news using artificial intelligence elements. *International Journal of Open Information Technologies*, *6* (12), 99-105.

Bio Note:

Alexandra M. Kasimova, Bachelor student, RUDN University named after Patrice Lumumba, Miklukho-Maklaya str., 6, Moscow, 117198, Russia. ORCID: 0009-0004-4193-9935. E-mail: 2241133074@pfur.ru.

Ekaterina S. Kozlovskaya, Associate Professor, Department of Russian Language No. 5, RUDN University named after Patrice Lumumba, Miklukho-Maklaya str., 6, Moscow, 117198, Russia. ORCID: 0000-0002-6308-725X. E-mail: kozlovskaya_es@pfur.ru.