

ყალბი ამბები ხელოვნური ინტელექტის ეპოქაში - საფრთხეები და მათი დაძლევის გზები

დავით ქუტიძე¹

თანამედროვე გამოწვევა - ხელოვნური ინტელექტით გენერირებული ყალბი ამბები

ხელოვნურმა ინტელექტმა ფუნდამენტურად შეცვალა ზოგადად საინფორმაციო გარემო და ამავდროულად, ყალბი ამბების შექმნის, გავრცელებისა და გამოვლენის პრაქტიკები. ყალბი ამბების ფენომენის მკვლევრები დღევანდელ რეალობას აღწერენ, როგორც „მნიშვნელოვან საფრთხეს მსოფლიო საინფორმაციო ეკოსისტემისთვის“.² ყალბი ამბების გავრცელებას, რაშიც, ძირითადად, მოიაზრებენ მისინფორმაციასა (ცრუ ინფორმაცია, რომელიც გავრცელებულია ბოროტი განზრახვის გარეშე) და დეზინფორმაციას (განზრახ შეთხზული ამბავი, რომელიც მიზნად ისახავს აუდიტორიის შეცდომაში შეყვანას), მნიშვნელოვნად უწყობს ხელს დახვეწილი გენერაციული ხელოვნური ინტელექტის სისტემები, დიდი ენობრივი მოდელები (Large Language Models – (LLM)) და ე. წ. გაყალბებული ვიდეოები (Deepfake).³

„დიდი ენობრივი მოდელები (LLM) არის მოწინავე ხელოვნური ინტელექტის სისტემები, რომლებიც იყენებს უზარმაზარი რაოდენობის მონაცემებსა და დახვეწილ ალგორითმებს ადამიანის ენის გასაგებად და შესაბამისი უკუკავშირის გენერირებისთვის“.⁴ LLM ტექსტზე დაფუძნებული დეზინფორმაციის გენერირებისთვის ეფექტურ ხელსაწყოს წარმოადგენს. კვლევები აჩვენებს, რომ LLM-ს შეუძლია შექმნას ძალიან დამაჯერებელი გაყალბებული ინფორმაცია, რომლის ტრადიციული მეთოდებით ამოცნობა დიდ სირთულეებთან არის დაკავშირებული.⁵ აღნიშნული მოდელები ქმნის უპრეცედენტო მასშტაბის შესაძლებლობას ყალბი სიხალეების, სოციალური მედიის პოსტებისა და მანიპულაციური ნარატივების

¹ კვლევითი ინსტიტუტი Gnomon Wise; e-mail: d.kutidze@ug.edu.ge

² M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

³ Ibid;

Park, S., Nan, X. Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. *AI & Soc* 41, 1501–1515 (2026). <https://doi.org/10.1007/s00146-025-02620-3>

⁴ ცხველიანი, გ. (2 თებერვალი, 2024). „რა არის დიდი ენობრივი მოდელი (LLM)“. Founders.ge. ხელმისაწვდომია: <https://bit.ly/4tVmELw>

⁵ Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

შექმნისთვის.⁶ Sun et al. (2024) თანახმად, LLM-ის მიერ გენერირებული ტექსტები, ლინგვისტური დახვეწილობისა და კონტექსტური თანმიმდევრულობის გამო, ინფორმაციის მომხმარებლებისთვის მნიშვნელოვნად ართულებს მათ ამოცნობას.⁷ აღნიშნულ მოდელებს შეუძლია წერის სტილის ადაპტირება, ფაქტობრივი ელემენტების ჩართვა სანდოობის გასაზრდელად და კონკრეტული აუდიტორიისა თუ პლატფორმისთვის მორგებული შინაარსის გენერირება.⁸ Yi et al. (2025) აღნიშნავს, რომ ხელოვნური ინტელექტის განვითარება, ერთი მხრივ, იძლევა ახალ ხელსაწყოებს ყალბი ამბების აღმოჩენისთვის, მაგრამ, ამის პარალელურად, ქმნის ახალ გამოწვევებსაც, რადგან მისი უნარი - შექმნას მაქსიმალურად „ადამიანური“ ტექსტები - სულ უფრო იზრდება. ვითარებას ისიც ართულებს, რომ LLM-ის გამოყენება შესაძლებელია არა მხოლოდ ცალკეული ყალბი სიახლის შესაქმნელად, არამედ ისეთი კოორდინირებული დეზინფორმაციული კამპანიების საწარმოებლადაც, რომლებიც ხასიათდება განმეორებადი, თანმიმდევრული ნარატივებითა და სოციალური ქსელების მრავალ პლატფორმაზე გავრცელების შესაძლებლობით.⁹

ხელოვნური ინტელექტისა და ყალბი ამბების კონტექსტში, კიდევ ერთი საყურადღებო საკითხია ე. წ. Deepfake ტექნოლოგია. „Deepfake არის ხელოვნურად გენერირებული მედია, რომელიც ციფრულადაა მანიპულირებული გარკვეული პირის აუდიო/ვიზუალური მონაცემების შესაცვლელად. “Deep” ტერმინში იგულისხმება ფოტო თუ ვიდეო მასალის გაყალბება „Deep learning“ მეთოდით, რომელიც მოიაზრებს მანქანური სწავლების ღრმა აღქმა-ათვისების მეთოდს ხელოვნური ნეირონული ქსელებით (ANN)“.¹⁰ კვლევების თანახმად, Deepfake წარმოადგენს განსაკუთრებით სახიფათო ფენომენს იმ თვალსაზრისით, რომ ადამიანები უფრო მეტად

⁶Jingyuan Yi, Zeqiu Xu, Tianyi Huang, and Peiyang Yu. 2025. Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. In Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security (GAIS '25). Association for Computing Machinery, New York, NY, USA, 87–93. <https://doi.org/10.1145/3728725.3728739>

⁷Sun, Yanshen, Jianfeng He, Limeng Cui, Shuo Lei and Chang-Tien Lu. “Exploring the Deceptive Power of LLM-Generated Fake News: A Study of Real-World Detection Challenges.” *ArXiv* abs/2403.18249 (2024).

⁸Fatimah, Raffharum, Auziah Mumtaz, Fauzan Muhammad Fahrezi and Diky Zakaria. “AI-Generated Misinformation: A Literature Review.” *Indonesian Journal of Artificial Intelligence and Data Mining* (2024).

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>

⁹Jingyuan Yi, Zeqiu Xu, Tianyi Huang, and Peiyang Yu. 2025. Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. In Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security (GAIS '25). Association for Computing Machinery, New York, NY, USA, 87–93. <https://doi.org/10.1145/3728725.3728739>

¹⁰Case. (3 თებერვალი, 2024). „რა არის Deepfake? განმარტება & საფრთხეები“. ხელმისაწვდომია: <https://globalcase.org/ge/blog/ra-aris-deepfake/>

ენდობიან ვიზუალურ ან აუდიო მტკიცებულებებს, ვიდრე, მაგალითად, ტექსტს. ამ ფონზე კი, Deepfake ქმნის ერთგვარ „ინფოკალიფს“ (Rajagopal et al. 2023), სადაც საკუთარი თვალით ნანახი ან მოსმენილი ინფორმაციაც კი აღარ ნიშნავს, რომ იგი ასი პროცენტით სარწმუნოა.¹¹ ამასთან, Deepfake-ების სწრაფი ტემპებით განვითარება უზრუნველყოფს იმას, რომ ისინი უფრო და უფრო რეალისტური ხდება. ეს ტექნოლოგიები სახის შეცვლის, ხმის კლონირებისა და სხულის ფორმების მიმსგავსების საშუალებასაც იძლევა, რაც დამაჯერებლად წარმოაჩენს კონკრეტული ინდივიდის ისეთ ქცევას ან ნათქვამს, რომელიც, სინამდვილეში, არასდროს მომხდარა. მსგავსი შესაძლებლობები სცდება წმინდა ტექნიკურ საკითხებს და დიდ გავლენას ახდენს მთლიანად საზოგადოებაზე. Deepfake საფრთხეს უქმნის ადამიანის პირად სივრცეს, აზიანებს პოლიტიკურ დისკურსს საჯარო პერსონების გაყალბებული განცხადებებით და ძირს უთხრის სანდო მედიასაშუალებების ავტორიტეტს საზოგადოების თვალში.¹²

ყალბი ამბების კონტექსტში მნიშვნელოვანია გენერაციული ხელოვნური ინტელექტის განხილვაც. „გენერაციული ხელოვნური ინტელექტი (Gen AI) წარმოადგენს ხელოვნური ინტელექტის ისეთ ტიპს, რომელიც ადამიანის ჩარევის გარეშე ქმნის შინაარსს - ტექსტს, გამოსახულებებს და მუსიკასაც კი. ეს ტექნოლოგია იყენებს რთულ ალგორითმებსა და მანქანური სწავლების მოდელებს არსებული მონაცემებიდან კანონზომიერებებისა და წესების დასამახსოვრებლად. შედეგად, იქმნება ახალი შინაარსი, რომელიც სტილითა და სტრუქტურით ორიგინალის მსგავსია“.¹³ თანამედროვე გენერაციული ხელოვნური ინტელექტის სისტემები სულ უფრო ეფექტიანად მოქმედებს მრავალ განზომილებაში, აერთიანებენ ტექსტს, სურათებს, აუდიო- და ვიდეომასალას რთული დეზინფორმაციული კამპანიების შესაქმნელად. GenAI-ის კვლევა ავლენს, ერთი შეხედვით, სარწმუნო, მაგრამ შეთხზული შინაარსის შექმნის შესაძლებლობებს, რომლებიც, მეტი დამაჯერებლობისთვის, მედიის მრავალ ტიპს აერთიანებს.¹⁴ ეს ინტეგრაცია დეზინფორმაციის აღმოჩენას უფრო რთულს ხდის, რადგან

¹¹Rajagopal, Tendral, Velayutham Chandrashekar, and Vignesh Ilango. 2023. "Unmasking the Deepfake Infocalypse: Debunking Manufactured Misinformation With a Prototype Model in the AI Era 'Seeing and Hearing, No Longer believing.'". *Journal of Communication and Management* 2 (04). India:230-37. <https://doi.org/10.58966/JCM2023243>.

¹²Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*, 5(2), 199–217. <https://doi.org/10.1080/23738871.2020.1797135>

¹³Polux Media. (31 ოქტომბერი, 2024). „გენერაციული ხელოვნური ინტელექტი (Gen AI) და მისი ეკონომიკური გავლენა“. ხელმისაწვდომია: <https://polux.ge/gen-ai-economics/>

¹⁴M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723

მოითხოვს მრავალი კომპონენტის თანადროულ ანალიზსა და მათ შორის შეუსაბამობების გამოვლენას.

როგორც აღვნიშნეთ, ხელოვნური ინტელექტით გენერირებულ ყალბ ამბებს ან მასშტაბურ დეზინფორმაციულ კამპანიებს სამიზნე სოციალურ გავლენის მოხდენის შესაძლებლობა აქვს. არაერთი კვლევის თანახმად, მსგავსი კამპანიები გამოყენებულია საარჩევნო პროცესების, ზოგადი პოლიტიკური დისკურსისა თუ სამოქალაქო ჩართულობის წინააღმდეგ.¹⁵ სხვადასხვა მნიშვნელოვანი მოვლენის შესწავლა ასახავს, თუ როგორ შეუძლია ხელოვნური ინტელექტით გენერირებულ დეზინფორმაციას, სწრაფად გავრცელოს პოლიტიკურად მოტივირებული, მანიპულირებული „ახალი ამბავი,“ რომელიც გავლენას ახდენს საზოგადოებრივ აზრსა და ქცევაზე. ხელოვნური ინტელექტის სისტემების შესაძლებლობა, მასშტაბურად წარმოქმნას მიზანმიმართული დეზინფორმაცია, იძლევა დახვეწილი გავლენის ოპერაციების ჩატარების საშუალებას, კონკრეტული დემოგრაფიული ჯგუფების ან გეოგრაფიული არეალის მიზანში ამოსაღებად. დემოკრატიულ სისტემებზე მიტანილი ეს იერიშები შესაბამის ნეგატიურ შედეგებს იწვევს. მკვლევრები აღნიშნავენ, რომ ხელოვნური ინტელექტით გენერირებულ ყალბ ამბებს შეუძლია ამომრჩეველთა აქტივობაზე ზეწოლა, პოლიტიკური პოლარიზაცია, საზოგადოებაში ინსტიტუტებისადმი ნდობის შერყევა და კრიტიკულ პოლიტიკურ საკითხებზე საზოგადოებრივი აზრით მანიპულირება.¹⁶

ისევე, როგორც ტრადიციული მეთოდებით გავრცელებულ დეზინფორმაციას, ხელოვნური ინტელექტით გენერირებულ ყალბ ნარატივებსაც აქვს ნეგატიური ეფექტი ზოგად საინფორმაციო ეკოსისტემასა და საზოგადოების ნდობაზე. ამ სფეროს მკვლევართა შემფოთების საგანს წარმოადგენს ინფორმაციული ეკოსისტემების მიმართ საზოგადოების ნდობის ეროზია. კვლევები აჩვენებს, თუ როგორ ქმნის ხელოვნური ინტელექტით გენერირებული ყალბი ამბების გავრცელება ფენომენს, რომელსაც მკვლევრები „ინფორმაციულ აშლილობას“ უწოდებენ, როცა

¹⁵Ibid.

Gaborit, P. (2024). A Sociopolitical Approach to Disinformation and AI: Concerns, Responses and Challenges. *Journal of Political Science and International Relations*, 7(4), 75-88. <https://doi.org/10.11648/j.ipsir.20240704.11>

¹⁶Danni Xu, Shaojing Fan, and Mohan Kankanhalli. 2023. Combating Misinformation in the Era of Generative AI Models. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. Association for Computing Machinery, New York, NY, USA, 9291–9298. <https://doi.org/10.1145/3581783.3612704>

Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*, 5(2), 199–217. <https://doi.org/10.1080/23738871.2020.1797135>

Gaborit, P. (2024). A Sociopolitical Approach to Disinformation and AI: Concerns, Responses and Challenges. *Journal of Political Science and International Relations*, 7(4), 75-88. <https://doi.org/10.11648/j.ipsir.20240704.11>

რიგითი მოქალაქეებისთვის სულ უფრო რთულდება ავტენტური შინაარსის გამოგონილისგან გარჩევა.¹⁷ შედეგად, შეიძლება ითქვას, „ბინძურდება“ საინფორმაციო ველი და რთულდება სანდო და არასანდო წყაროების ერთმანეთისგან გამიჯვნა. ამ თემაზე ჩატარებული კვლევები ავლენს შემაშფოთებელ პარადოქსს. კერძოდ, მიუხედავად იმისა, რომ დეზინფორმაციის შესახებ საზოგადოების ცნობიერება გაიზარდა, ყალბი ამბების ამოცნობისა და მათ უკან მდგომი პროცესების გაგების უნარი კვლავ შეზღუდულია.¹⁸ საბოლოოდ, ეს ნეგატიური გავლენა ინსტიტუციურ ნდობაზეც ვრცელდება. კვლევები მიუთითებს, რომ ხელოვნური ინტელექტით გენერირებული ყალბი ამბები ამცირებს საზოგადოების ნდობას მედიაორგანიზაციების, სამეცნიერო ინსტიტუტებისა და სამთავრობო უწყებების მიმართ.¹⁹

კოლექტიური ზემოქმედების გარდა, ხელოვნური ინტელექტით გენერირებული ყალბი ამბები სერიოზულ ინდივიდუალურ რისკებსაც წარმოშობს. ე. წ. Deepfake-ები ხელს უწყობს პირად მონაცემებთან დაკავშირებული ყალბი შინაარსის უნებართვოდ შექმნას, რაც ხშირად არღვევს პირადი ცხოვრების კონფიდენციალურობას. კვლევები აღწერს ე. წ. სინთეტიკური მედიის გამოყენების შემთხვევებს შევიწროების, რეპუტაციის შელახვისა და თაღლითობისთვის.²⁰ აგრეთვე, ხელოვნური ინტელექტით გენერირებული ყალბი ამბების ზემოქმედებამ შეიძლება გამოიწვიოს ინდივიდის დაბნეულობა, შფოთვა და ცინიზმი ინფორმაციის წყაროების მიმართ.²¹

¹⁷Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

Salubi, O. (2025), Artificial Intelligence, Misinformation, and Libraries: A New Frontier for Information Professionals. *Proceedings of the Association for Information Science and Technology*, 62: 572-581. <https://doi.org/10.1002/pr2.1279>

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>

¹⁸Gaborit, P. (2024). A Sociopolitical Approach to Disinformation and AI: Concerns, Responses and Challenges. *Journal of Political Science and International Relations*, 7(4), 75-88. <https://doi.org/10.11648/j.ijpsir.20240704.11>

¹⁹M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>

²⁰M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

²¹Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

ხელოვნური ინტელექტით გენერირებული ყალბი ამბების აღმოჩენის გზები

ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ ტექნიკური კონტროლების უმეტესობის საფუძველს მანქანური სწავლების მიდგომები წარმოადგენს. ეს მიდგომები მოიცავს ენობრივი ნიმუშების (სტილისტური მახასიათებლები, ტექსტის ლოგიკური თანწყობის ანალიზი, ფაქტობრივი თანმიმდევრულობა), ვიზუალური არტეფაქტების (ფოტოებში გენერირების კვალის ძიება, ვიდეოში არსებული შეუსაბამობები) და ქცევითი სიგნალების (გავრცელების პატერნები, არაავთენტური კოორდინირებული ქცევა) ანალიზს.²² თუმცა აღნიშნულ მიდგომებს აქვს თავისი ლიმიტები, რადგან აღმოჩენის სისტემების გაუმჯობესებასთან ერთად, ადაპტირდება გენერაციის მოდელებიც აღმოჩენის თავიდან ასაცილებლად, რაც ქმნის ერთგვარ დაუსრულებელ რბოლას. იმის გათვალისწინებით, რომ ხელოვნური ინტელექტით გენერირებული ყალბი ამბები ხშირად მოიცავს მრავალ განზომილებას, მკვლევრები გვთავაზობენ აღმოჩენის მულტიმოდალური მიდგომებს, რომლებიც ერთდროულად ანალიზებს ტექსტს, სურათებს, აუდიო და ვიდეო მასალებს. ეს სისტემები ადგენს შეუსაბამობებს მოდალობებს შორის, მაგალითად, შეუსაბამობებს ტუჩების მოძრაობებსა და ხმას შორის ვიდეოში, ან წინააღმდეგობებს ტექსტსა და სურათებს შორის.²³ ხელოვნური ინტელექტის გენერირებული შინაარსის აღმოსაჩენად მნიშვნელოვანი ხელსაწყოა ე.წ. ციფრული მარკირება, რომელიც გულისხმობს ფოტოებისა თუ ვიდეოების გენერირების პროცესში როგორც ხილული, ისე უხილავი ნიშნების ჩანერგვას. ეს პროცესი ხელოვნური ინტელექტის პლატფორმებისგან ისეთი სპეციფიკური ნიშნების გამოყენებას მოითხოვს, რომლებიც სხვადასხვა ტიპის მანიპულაციებს (მაგ. შეკუმშვა, ამოჭრა, ფილტრაცია და ა.შ.) გაუძლებს და შემოწმების სისტემებისთვის კვლავ ამოსაცნობი იქნება.

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>

²²Danni Xu, Shaojing Fan, and Mohan Kankanhalli. 2023. Combating Misinformation in the Era of Generative AI Models. In Proceedings of the 31st ACM International Conference on Multimedia (MM '23). Association for Computing Machinery, New York, NY, USA, 9291–9298. <https://doi.org/10.1145/3581783.3612704>

Aos Mulahuwaish, Basheer Qolomany, Kevin Gyorick, Jacques Bou Abdo, Mohammed Aledhari, Junaid Qadir, Kathleen Carley, Ala Al-Fuqaha, A survey of social cybersecurity: Techniques for attack detection, evaluations, challenges, and future prospects, Computers in Human Behavior Reports, Volume 18, 2025, 100668, ISSN 2451-9588, <https://doi.org/10.1016/j.chbr.2025.100668>

²³M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," 2023 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

ლიტერატურა აღწერს ხელოვნური ინტელექტით გენერირებულ ყალბ ამაბებზე პოლიტიკის მიდგომებსა და მარეგულირებელ რეაგირებას, რომელთა შორის განსაკუთრებული ყურადღება ეთმობა ევროპულ ინიციატივებს. Gaborit-ი ხაზს უსვამს ევროკავშირის ხელოვნური ინტელექტის აქტს²⁴, როგორც მნიშვნელოვან მარეგულირებელ ჩარჩოს, რომელიც ეხება ხელოვნური ინტელექტით გენერირებულ დეზინფორმაციას. ეს საკანონმდებლო აქტი ადგენს გამჭვირვალობის, რისკის შეფასებისა და ანგარიშვალდებულების მოთხოვნებს იმ ხელოვნური ინტელექტის სისტემებისთვის, რომლებსაც შეუძლია დეზინფორმაციის გენერირება ან გავრცელება.²⁵ Shoaib et al. (2023)²⁶ ემხრობა გენერაციული ხელოვნური ინტელექტის ეთიკური გამოყენების შესახებ გლობალური კონსენსუსის მიღწევას და კოორდინირებული პოლიტიკური ინიციატივების განხორციელებას სხვადასხვა იურისდიქციაში. ტრანსნაციონალური დეზინფორმაციული კამპანიების გამოწვევა მოითხოვს ჰარმონიზებულ მარეგულირებელ მიდგომებს, რომლებსაც შეუძლია მრავალი იურისდიქციიდან წარმოშობილი კონტენტის დაფარვა. კვლევები ასევე საუბრობს სექტორულ რეგულაციებზე, რომელთა ფარგლებშიც შესწავლილია სოციალური მედიის პლატფორმების პოლიტიკა, რომელიც მოითხოვს გამჭვირვალობას კონტენტის მოდერაციაში და თანამშრომლობას ფაქტების გადამმოწმებელ (Fact-Checking) ორგანიზაციებთან. თუმცა Gaborit-ი (2024) აღნიშნავს, რომ AI „დეზინფორმაციაზე რეაგირება ჯერ კიდევ ადრეულ ეტაპზეა“, რაც მიუთითებს მარეგულირებელი ჩარჩოს ეფექტურობის მნიშვნელოვან ხარვეზებზე.

აქვე აღსანიშნავია, რომ ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ ხედვები და პოლიტიკა აქვს სოციალური ქსელის პლატფორმებსაც. თუმცა აქაც ვლინდება ეფექტურობასთან დაკავშირებული გამოწვევები. სოციალური ქსელების პლატფორმების ინიციატივებიდან, აღსანიშნავია კონტენტის ეტიკეტირების (ხელოვნური ინტელექტით გენერირებული კონტენტის მონიშვნა), ალგორითმზე დაფუძნებული მიდგომების (საეჭვო დეზინფორმაციის ხილვადობის შემცირება) და ანგარიშების შეზღუდვების (ყალბი ამბების განმეორებით გაზიარების შეზღუდვა) მცდელობები. თუმცა არსებობს მოსაზრებები, რომ ინდუსტრიის ნებაყოფლობით ინიციატივებს აკლია აღსრულების მექანიზმები და ანგარიშვალდებულება. გარდა ამისა, კონტენტის მოდერაციას აქვს კიდევ ერთი გამოწვევა -

²⁴The EU Artificial Intelligence Act. Via link: <https://artificialintelligenceact.eu/>

²⁵Gaborit, P. (2024). A Sociopolitical Approach to Disinformation and AI: Concerns, Responses and Challenges. *Journal of Political Science and International Relations*, 7(4), 75-88. <https://doi.org/10.11648/j.jpsir.20240704.11>

²⁶M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

საფრთხე სიტყვისა და გამოხატვის თავისუფლებისთვის. კერძოდ, აგრესიულმა ანტიდევინფორმაციულმა პოლიტიკამ შეიძლება ჩაახშოს თავისუფალი სიტყვა ან შექმნას ცენზურის რისკები, რადგან ხშირად ზღვარი მანიპულირებულ ინფორმაციასა და გამოხატვის თავისუფლებას შორის ძალიან ბუნდოვანია.

ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ, ბუნებრივია, არ არსებობს ერთი კონკრეტული მექანიზმი, რომელიც პრობლემას სრულად მოაგვარებს. ამდენად, მკვლევრები ხაზს უსვამენ სექტორებს შორის თანამშრომლობის აუცილებლობას, რომელშიც მონაწილეობენ ტექნოლოგიური კომპანიები, სამთავრობო უწყებები, აკადემიური ინსტიტუტები, სამოქალაქო საზოგადოების ორგანიზაციები და მედიასაშუალებები.²⁷ წარმატებული თანამშრომლობის მაგალითები შეიძლება იყოს საჯარო და კერძო პარტნიორობა საფრთხეების შესახებ ინფორმაციის გაზიარებით, აკადემიისა და AI ინდუსტრიის თანამშრომლობა ყალბი ამბების აღმოჩენის ტექნოლოგიების განვითარებისთვის და ინიციატივები ციფრული წიგნიერების გასავრცელებლად.

ზემოთ ჩამოთვლილი კონტრზომებიდან, განსაკუთრებულ ყურადღებას იპყრობს ე.წ. ციფრული წიგნიერება, რომელიც გულისხმობს ინდივიდების იმ ცოდნითა და უნარებით აღჭურვას, რომლებიც აუცილებელია თანამედროვე ციფრულ სამყაროში არსებობისთვის. ბუნებრივია, საზოგადოების დიდი ნაწილის შესაბამისი უნარებით აღჭურვა საკმაოდ დიდ დროსა და რესურსს მოითხოვს, თუმცა, ამავდროულად, შესაძლოა ყველაზე ეფექტური და უსაფრთხო გზა იყოს დევინფორმაციის წინააღმდეგ ბრძოლაში.

ინდივიდზე ორიენტირებული მიდგომები და ციფრული წიგნიერება

კვლევების თანახმად, ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ ბრძოლის ერთ-ერთი მთავარი იარაღი ციფრული და მედიაწიგნიერებაა, რაც ინდივიდების კრიტიკული აზროვნების განვითარებასა და მათი იმ უნარებით აღჭურვას გულისხმობს,

²⁷Gilbert, C., & Gilbert, M. A. (2024). Leveraging Artificial Intelligence (AI) by a Strategic Defense against Deepfakes and Digital Misinformation. *International Journal of Scientific Research and Modern Technology*, 3(11), 62–78. <https://doi.org/10.38124/ijsrmt.v3i11.76>.

Danni Xu, Shaojing Fan, and Mohan Kankanhalli. 2023. Combating Misinformation in the Era of Generative AI Models. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. Association for Computing Machinery, New York, NY, USA, 9291–9298. <https://doi.org/10.1145/3581783.3612704>.

რომლებიც დეზინფორმაციის იდენტიფიცირებაში დაეხმარებათ.²⁸ მეტიც, მსგავსი ძალისხმევა არა მხოლოდ სიყალბის იდენტიფიცირების შესაძლებლობას აძლევს ადამიანს, არამედ მასთან შეწინააღმდეგების უნარსაც. თავის მხრივ, ციფრული წიგნიერება რამდენიმე მნიშვნელოვან კომპონენტს მოიცავს, ესენია: ხელოვნური ინტელექტის შესაძლებლობებისა და ლიმიტების ცოდნა; ყველაზე გავრცელებული მანიპულაციური ტექნიკების ამოცნობა; ინფორმაციის წყაროების გადამოწმება; ახალი ამბის შინაარსის სანდოობის შეფასება და ე.წ. ჯანსაღი სკეპტიციზმი (რომელიც არ უნდა გადაიზარდოს ცინიზმსა და მედიის მიმართ სრულ უნდობლობაში) ნებისმიერი ახალი ამბის მიმართ.²⁹ ამასთან, ციფრული წიგნიერების ხელშემწყობი კამპანიები არ უნდა იყოს ერთჯერადი ხასიათის და მუდმივად უნდა ვითარდებოდეს ხელოვნური ინტელექტის განვითარებასთან ერთად.³⁰

ამდენად, მიუხედავად იმისა, რომ ინსტიტუციური რეაგირება, ტექნიკური გადაწყვეტები და საკანონმდებლო რეგულაციები იძლევა ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ საბრძოლველ მნიშვნელოვან ინფრასტრუქტურას, ინფორმაციულ ეკოსისტემებში ინდივიდები კვლავაც თავდაცვის პირველ ხაზს წარმოადგენენ. შესაბამისად, ქვემოთ სწორედ იმ კონკრეტულ პრაქტიკულ ხერხებს განვიხილავთ, რომლებიც ადამიანებს ციფრულ სამყაროში ოპერირებისთვის გამოადგებათ.

AI დეზინფორმაციის აღმოჩენის პრაქტიკული ხერხები

- მანიშნებელი არტეფაქტები - ესენია მანიპულაციის შესამჩნევი კვალი აუდიო, ვიდეო ან ფოტომასალაში. ყურადღება უნდა მიექცეს უჩვეულო განათებას, ჩრდილებს, უჩვეულო ფერებსა და საგნების ფორმებსაც კი, რომლებიც ხშირად არაბუნებრივი

²⁸Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

Salubi, O. (2025), Artificial Intelligence, Misinformation, and Libraries: A New Frontier for Information Professionals. *Proceedings of the Association for Information Science and Technology*, 62: 572-581. <https://doi.org/10.1002/pra2.1279>.

Singh, Kawaljit & Kumar, Harish. (2025). Futuristic Media Information Literacy to Counter AI Generative Deepfake Media Content and its Implication. *Jharkhand Journal of Development and Management Studies*. 23. <https://www.xiss.ac.in/JJDMs/current/article/view/331>.

²⁹Salubi, O. (2025), Artificial Intelligence, Misinformation, and Libraries: A New Frontier for Information Professionals. *Proceedings of the Association for Information Science and Technology*, 62: 572-581. <https://doi.org/10.1002/pra2.1279>.

³⁰Jiang, Bohan, Zhen Tan, Ayushi Nirmal, and Huan Liu. 2024. Disinformation Detection: An Evolving Challenge in the Age of LLMs., 427–435. doi:10.1137/1.9781611978032.50.

მოყვანილობისაა. ვიდეოკონტენტში განსაკუთრებული ყურადღება უნდა მიექცეს სახეებსა და ფონს შორის საზღვრებს, თვალეებში ან სათვალეებში ანარეკლებსა და კადრის სხვადასხვა ელემენტში განათების თანმიმდევრულობას.

- აღქმის უცნაურობები - ვლინდება შეუსაბამობებით, რომლებიც განსაკუთრებით Deepfake ვიდეოებს ახლავს თან. ესენია: სახის არაბუნებრივი გამომეტყველება, ტუჩების სინქრონიზაციის შეცდომები, სადაც პირის მოძრაობები ზუსტად არ ემთხვევა ხმას, ასევე თვალეების უჩვეულო ხამხამი, თავის არათანმიმდევრული მოძრაობები და ა.შ.
- სემანტიკური შეუსაბამობები - აქ განსაკუთრებული ყურადღება უნდა მიექცეს მასალის შინაარსს, რადგან ხელოვნური ინტელექტით გენერირებულ მასალაში ხშირია ფაქტობრივი წინააღმდეგობები, მოვლენების არაზუსტი თანმიმდევრობით წარმოდგენა დროით ქრილში, ლოგიკური ხარვეზები და ისეთი მტკიცებები, რომლებიც ზოგჯერ საყოველთაოდ გავრცელებულ ცოდნასაც კი ეწინააღმდეგება. ასევე, ხელოვნური ინტელექტით გენერირებულ ტექსტებში შეიძლება შეგვხვდეს ზედმეტად ხშირად განმეორებადი ფრაზები, ძალიან ზოგადი, დეტალების გარეშე გადმოცემული ინფორმაცია და ისეთი ენობრივი შეცდომები, რომლებიც, როგორც წესი, ადამიანებისთვის ნაკლებად არის დამახასიათებელი.
- ემოციური ტონი და დარწმუნების მიზნით განმეორებადი სიტყვები - კვლევები მიუთითებს, რომ გენერაციული მოდელები ხშირად გამოიყენება ისეთი მანიპულაციური ნარატივების შესაქმნელად, რომლებიც განსაკუთრებულად ეხმიანება იდენტობასა და ინდივიდის ემოციებს. ასევე, საზოგადოების ნაწილში უკვე კარგად გამჯდარ სტერეოტიპებს. ინფორმაცია, რომელიც იწვევს ინტენსიურ ემოციურ რეაქციებს - აღშფოთებას, შიშს, აღელვებას - განსაკუთრებით კარგად შესწავლას საჭიროებს.³¹

³¹Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

Jingyuan Yi, Zeqiu Xu, Tianyi Huang, and Peiyang Yu. 2025. Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. In Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security (GAIS '25). Association for Computing Machinery, New York, NY, USA, 87–93. <https://doi.org/10.1145/3728725.3728739>.

Sun, Yanshen, Jianfeng He, Limeng Cui, Shuo Lei and Chang-Tien Lu. "Exploring the Deceptive Power of LLM-Generated Fake News: A Study of Real-World Detection Challenges." *ArXiv abs/2403.18249* (2024).

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In Proceedings of the

კოგნიტური მდგრადობის ტექნიკები

ახალი ინფორმაციის ტექნიკური ანალიზის გარდა, არსებობს ფსიქოლოგიური და ქცევითი სტრატეგიები, რომლებიც ხელოვნური ინტელექტით გენერირებული დეზინფორმაციის წინააღმდეგ ბრძოლაში გვეხმარება.

- პაუზა ახალი სიახლის გაზიარებამდე - შეიძლება ითქვას, ეს ყველაზე მნიშვნელოვანი ინდივიდუალური ძალისხმევაა ყალბი ამბების წინააღმდეგ. კვლევები აჩვენებს, რომ მოულოდნელი ან ემოციური შინაარსის დაუყოვნებლივ გაზიარების ნაცვლად მცირე (დაახლოებით ათწამიანი) პაუზის გაკეთება, რეალურად ამცირებს დეზინფორმაციის ეფექტებს. ამდენად, როგორი ემოციური თუ სკანდალურიც უნდა იყოს ესა თუ ის მასალა, მნიშვნელოვანია მისი მყისიერად გაზიარების იმპულსის დათრგუნვა, რაც, საბოლოო ჯამში, არღვევს დეზინფორმაციის სწრაფი გავრცელების ტენდენციას. სწრაფი გავრცელება კი სწორედ ისაა, რაც ვირუსული დეზინფორმაციის კამპანიებს ეფექტურს ხდის.
- ალტერნატიული წყაროები - ახალი ამბის დაჯერებამდე ან ჩვენი წინასწარი განწყობების დადასტურებამდე, მნიშვნელოვანია, რომ ვეწვიოთ რამდენიმე დამოუკიდებელ წყაროს (ვებგვერდს ან სოციალური ქსელის შესაბამის გვერდებს), ან მივმართოთ ფაქტების გადამმოწმებელ ავტორიტეტულ ორგანიზაციებს. ეს მეთოდი საშუალებას იძლევა, ინფორმაცია რამდენიმე სხვადასხვა წყაროდან მივიღოთ და შესაბამისად, განსხვავებული პერსპექტივებიდან შევხედოთ მას.
- რა ინტერესი აქვს ინფორმაციის წყაროს? - უნდა დაისვას შემდეგი კითხვები: ვისთვის მოაქვს სარგებელი ამ სიახლეს და აქვს, თუ - არა მის გამავრცელებელ წყაროს სანდოობა? ვინ შექმნა ეს მასალა? რა არის მათი მოტივაცია? აქვთ, თუ - არა მათ ექსპერტიზა ან ავტორიტეტი ამ თემაზე? იყვნენ, თუ - არა ისინი სანდო წარსულში?
- კოგნიტური მიკერძობების გამორიცხვა - ხშირ შემთხვევაში, ჩვენი წინასწარი განწყობები და კონკრეტული თემის მიმართ სუბიექტური დამოკიდებულებები გვაიძულებს, ნაკლებად ვიფიქროთ ამა თუ იმ სიახლის სანდოობაზე და ემოციური მსჯელობით, ასევე ჩვენი სურვილების/რწმენების დადასტურების მიზეზით, ნაკლებად კრიტიკულნი ვიყოთ სავარაუდო მანიპულაციის მიმართ. კვლევები ადასტურებს, რომ ჩვენი სუბიექტური განწყობების კარგად გააზრება და მათ შესახებ მიღებული სიახლეების განსაკუთრებით

2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>.

ფრთხილი ანალიზი მნიშვნელოვნად ამცირებს ინდივიდის შეცდომაში შეყვანის შესაძლებლობას.

- „ვარჯიში“ ციფრულ წიგნიერებაში, როგორც ციფრული უნარების განვითარების უწყვეტი პროცესი - წყაროების კრიტიკული ანალიზისა და ხელოვნური ინტელექტის შესაძლებლობების გაგების კომპეტენციების განვითარება რთულ ინფორმაციულ გარემოში უფრო ეფექტურ ნავიგაციას უზრუნველყოფს. საგანმანათლებლო დაწესებულებები და არა მხოლოდ, სულ უფრო ხშირად გვთავაზობენ ფართოდ ხელმისაწვდომ რესურსებს ამ უნარების განვითარებისთვის.³²

პლატფორმების მიერ მოწოდებული ტექნიკური ხელსაწყოების გამოყენება

- დეზინფორმაციის აღმოჩენის ავტომატური სისტემები - ეს სისტემები იყენებს მანქანურ სწავლებას იმისთვის, რომ ავტომატურ რეჟიმში მონიშნოს მანიპულირებული აუდიო, ვიდეო და ბეჭდური მასალები (მაგალითად, meta-ს პლატფორმებზე კონკრეტულ ვიდეოებზე შეიძლება შეგვხვდეს წარწერა AI info). პლატფორმები სულ უფრო ხშირად მიმართავენ მსგავს ხერხებს, ცდილობენ მონიშნონ საექვო ინფორმაცია და თან დაურთონ ფაქტების გადამმოწმებელი ავტორიტეტული ორგანიზაციების განმარტებები. ამდენად, მნიშვნელოვანია, მომხმარებლები კარგად გაეცნონ ამ ფუნქციებს მათ მიერ გამოყენებულ პლატფორმებზე და ყურადღება მიაქციონ შესაბამის გაფრთხილებებს. საწყის ეტაპზე მსგავსი მიდგომა ინდივიდს ყალბი ინფორმაციის შემჩნევაში ეხმარება. თუმცა ისიც გასათვალისწინებელია, რომ ამ გზით აბსოლუტურად ყველაფრის

³² Zahra, Yazdani. "Understanding and Mitigating AI-Generated Hoax Information." *Jurnal Sistem Informasi dan Teknik Informatika (JAFOTIK)* (2025).

M. R. Shoaib, Z. Wang, M. T. Ahvanooy and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 436, 1–20. <https://doi.org/10.1145/3544548.3581318>.

Salubi, O. (2025), Artificial Intelligence, Misinformation, and Libraries: A New Frontier for Information Professionals. Proceedings of the Association for Information Science and Technology, 62: 572-581. <https://doi.org/10.1002/pra2.1279>.

Singh, Kawaljit & Kumar, Harish. (2025). Futuristic Media Information Literacy to Counter AI Generative Deepfake Media Content and its Implication. *Jharkhand Journal of Development and Management Studies*. 23. <https://www.xiss.ac.in/JJDMs/current/article/view/331>.

გადამოწმება შეუძლებელია და არსებობს მატალი ხარისხის დეზინფორმაციული მასალები, რომელთა იდენტიფიცირებაც ავტომატურ სისტემებს ჯერ კიდევ უჭირთ.

- საგანმანათლებლო ორგანიზაციები და ფაქტების გადამოწმების სერვისები - ეს სერვისები მომხმარებლებს ასწავლის, თუ როგორ შეაფასონ ახალი ამბის სანდოობა და სთავაზობს სხვადასხვა ინსტრუმენტს მასალის გადასამოწმებლად. კვლევები აჩვენებს, რომ სანდო შუამავლებთან კონსულტაცია, როგორცაა ინფორმაციის გადამოწმების ექსპერტიზის მქონე ინსტიტუტები, ამცირებს ინდივიდების შეცდომაში შეყვანის მაჩვენებელს და ხელს უშლის დეზინფორმაციის უნებლიე გავრცელებას.
- მულტიმოდალური გადამოწმება - გულისხმობს ტექსტის, ვიდეოსა და სხვადასხვა კონტექსტური სიგნალის თანადროულ გადამოწმებას. საეჭვო მასალის შეფასებისას, უნდა შემოწმდეს, გადმოსცემს, თუ - არა სხვადასხვა მოდალობა (ტექსტი, ხმა, ვიზუალი) თანმიმდევრულ ამბავს. მაგალითად, ემთხვევა, თუ - არა ტექსტში მოცემული მტკიცებები სურათებში გადმოცემულ ინფორმაციას? ემთხვევა, თუ - არა კონკრეტული ხმები ვიდეოში არსებულ ვიზუალურ ელემენტებს? მსგავსი ჯვარედინი ანალიზის მეთოდი ხშირად ავლენს მანიპულირებულ ინფორმაციას.³³

შეჯამება

სხვადასხვა აკადემიური კვლევის შედეგები აჩვენებს, რომ ხელოვნური ინტელექტის სწრაფმა განვითარებამ რადიკალურად შეცვალა დეზინფორმაციის ბუნება და მასშტაბი, რის შედეგადაც ყალბი ამბები უფრო დამაჯერებელი, პერსონალიზებული და მასობრივად გავრცელებადი გახდა. დიდი ენობრივი მოდელები, გენერაციული AI და Deepfake ტექნოლოგიები ქმნის ისეთ საინფორმაციო გარემოს, რომელშიც ფაქტისა და სიყალბის ერთმანეთისგან გარჩევა

³³M. R. Shoaib, Z. Wang, M. T. Ahvanoey and J. Zhao, "Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models," *2023 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt, 2023, pp. 1-7, DOI: 10.1109/ICCA59364.2023.10401723.

Salubi, O. (2025), Artificial Intelligence, Misinformation, and Libraries: A New Frontier for Information Professionals. *Proceedings of the Association for Information Science and Technology*, 62: 572-581. <https://doi.org/10.1002/pra2.1279>.

Jingyuan Yi, Zeqiu Xu, Tianyi Huang, and Peiyang Yu. 2025. Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. In *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security (GAIS '25)*. Association for Computing Machinery, New York, NY, USA, 87–93. <https://doi.org/10.1145/3728725.3728739>.

Singh, Kawaljit & Kumar, Harish. (2025). Futuristic Media Information Literacy to Counter AI Generative Deepfake Media Content and its Implication. *Jharkhand Journal of Development and Management Studies*. 23. <https://www.xiss.ac.in/JJDMSC/current/article/view/331>.

მნიშვნელოვნად გართულებულია. ეს პროცესი აზიანებს როგორც საზოგადოებრივ დონეზე ინფორმაციულ ეკოსისტემასა და ინსტიტუტებისადმი ნდობას, ისე ინდივიდის ფსიქოლოგიურ მდგომარეობასა და ინფორმაციის აღქმის უნარს.

ტექნიკური თუ ინსტიტუციური პასუხების გარდა, შეიძლება ითქვას, ხელოვნური ინტელექტით გენერირებული ყალბი ამბების წინააღმდეგ ბრძოლაში განსაკუთრებული მნიშვნელობა ენიჭებათ ინდივიდებსა და მათზე ორიენტირებულ მიდგომებს, რომლებიც ადამიანებს აქცევს ინფორმაციის პასუხისმგებლიან მომხმარებლებად და არა მხოლოდ პასიურ მიმღებებად. ამ კონტექსტში, ციფრული წიგნიერება შეიძლება იყოს ყველაზე მდგრადი და ეფექტური პასუხი დეზინფორმაციის წინააღმდეგ. იგი მოიცავს არა მხოლოდ ტექნიკურ უნარებს, არამედ კრიტიკული აზროვნების განვითარებას, ინფორმაციის წყაროების შეფასების უნარსა და ხელოვნური ინტელექტის შესაძლებლობებისა თუ შეზღუდვების გააზრებას. ციფრული წიგნიერების მქონე ინდივიდი უკეთ ამოიცნობს მანიპულაციურ ნარატივებს, აფასებს შინაარსის სანდოობას და ნაკლებად ექვემდებარება ემოციურ მანიპულაციებს. ამ ყველაფრის მისაღწევად სფეროს მკვლევრები გამოყოფენ რამდენიმე ძირითად მეთოდს, რომელთა დახმარებითაც ტრადიციული თუ ალტერნატიული მედიის მომხმარებლები უფრო მომზადებულნი შეხვდებიან ხელოვნური ინტელექტით გენერირებულ ყალბ ამბებს. ესენია: ინფორმაციის გაზიარებამდე პაუზის გაკეთება, ალტერნატიული წყაროების გადამოწმება, ინფორმაციის გამავრცელებლის მოტივაციის გაანალიზება და საკუთარი კოგნიტური მიკროძოვებების გაცნობიერება. ასევე მნიშვნელოვანია მულტიმოდალური ანალიზის უნარი - ტექსტის, ვიზუალისა და ხმის თანხვედრის შემოწმება - რაც განსაკუთრებით აქტუალურია Deepfake-ების ეპოქაში. ეს მიდგომები ზრდის ე.წ. „კოგნიტურ მდგრადობას“, რაც ინდივიდს საშუალებას აძლევს, უკეთ გაუმკლავდეს მანიპულაციურ ინფორმაციას.

საბოლოოდ, ხელოვნური ინტელექტის ეპოქაში დეზინფორმაციის წინააღმდეგ ბრძოლა მოითხოვს კომპლექსურ, მრავალსექტორულ მიდგომას, რომლის მიხედვითაც, ტექნოლოგიური და ინსტიტუციური პასუხები უნდა შეივსოს ინდივიდის გაძლიერებით. სწორედ ციფრული წიგნიერება და ინდივიდის აქტიური ჩართულობა წარმოადგენს იმ „პირველ ხაზს“, რომელიც განსაზღვრავს, რამდენად მდგრადი იქნება საზოგადოება თანამედროვე საინფორმაციო გამოწვევების მიმართ.