International Journal of Social Sciences

Caucasus International University
Volume 6, Issue 3

Journal homepage: http://journal.ciu.edu.ge/

DOI: https://doi.org/10.55367/LOVB4005



*Conference proceedings – **International scientific conference "Contemporary Challenges in International Relations, International Law and Media"** organised by Caucasus International University*
Held on March 12-13, 2026 in Tbilisi

Artificial Intelligence, „Deepfakes" and International Information Warfare

George Khatidze [a1]

[a] PhD in Law, Professor
Caucasus International University

ARTICLE INFO

Keywords:

Artificial Intelligence

Deepfakes

International Information Warfare

ABSTRACT

The growing development of artificial intelligence affects our daily life. Its effect is multifaceted and characterized by positive as well as negative trends. The development of artificial intelligence creates new opportunities while also raising certain challenges. One of the negative practices of "deepfakes" is related to international information warfare. Therefore, regulation by means of legal and non-legal mechanisms of artificial intelligence and "deepfakes", which are based on its advanced technologies, is needed. The cross-border nature of "deepfakes" poses certain challenges in terms of their legal regulation. Another difficulty is that "deepfakes" are not criminalized. At the same time, it is difficult to identify and track down the creators and distributors of "deepfakes". The lack of a single, universally accepted definition of "deepfake" also presents a legal challenge. To overcome these challenges, both national legislation and various international instruments between states have an important role to play. Particular importance has the EU's 2024 "Artificial Intelligence Act", which imposes transparency requirements for certain content generated by artificial intelligence, including "deepfakes". However, it should be noted that the act is effective only throughout the European Union. Along with legal mechanisms, non-legal means are equally important. It can be noted that non-legal remedies are somewhat complementary in nature. The so-called "Platform Self-Regulation", the involvement of various technological companies, their coordination, media literacy, and raising awareness

[1] Corresponding author.
E-mail addresses: george.khatidze@ciu.edu.ge (G. Khatidze)
ORCID: https://orcid.org/0009-0006-6689-8505

Received: 11/04/2026 Revised: 09/07/2026 Accepted: 20/07/2026 Published: 26/11/2026
E-ISSN 2720-8036 / © 2026 The Author(s). Published in The International Journal of Social Sciences by Caucasus International University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

in society about the essence of artificial intelligence, its nature, and the "deepfakes" based on its technologies have an important role. However, non-legal mechanisms do not have binding force. Overall, it can be stated that the mechanisms of regulation of "deepfakes" are somewhat fragmented, with the lack of unified nature.

ხელოვნური ინტელექტი, „დიფფეიკები“ და საერთაშორისო საინფორმაციო ომი

გიორგი ხატიძე [a2]

[a] სამართლის დოქტორი, პროფესორი
კავკასიის საერთაშორისო უნივერსიტეტი

სტატიის შესახებ

საკვანძო სიტყვები:

ხელოვნური ინტელექტი

„დიფფეიკები“

საერთაშორისო საინფორმაციო ომი

აბსტრაქტი

ხელოვნური ინტელექტის განვითარება, რომელიც მზარდია, გავლენას ახდენს ადამიანთა ყოველდღიურ ცხოვრებაზე. მისი ეფექტი მრავალმხრივია და ხასიათდება როგორც დადებითი, ისე უარყოფითი ტენდენციებით. ხელოვნური ინტელექტის განვითარება ქმნის ახალ შესაძლებლობებს და, ამავდროულად, ბადებს გამოწვევებს. „დიფფეიკების“ გამოყენების ერთ-ერთი უარყოფითი პრაქტიკა საერთაშორისო საინფორმაციო ომს უკავშირდება. შესაბამისად, როგორც თავად ხელოვნური ინტელექტის, ისე მის მოწინავე ტექნოლოგიებზე დამყარებული „დიფფეიკების“ როგორც სამართლებრივი, ისე არასამართლებრივი გზებითა და საშუალებებით მოწესრიგება და რეგულირება საჭიროებას წარმოადგენს. სამართლებრივი მოწესრიგების კუთხით გამოწვევაა „დიფფეიკების“ ტრანსსასაზღვრო ბუნება. სირთულეს წარმოადგენს ასევე ის, რომ „დიფფეიკები“ არ არის კრიმინალიზებული. ამავდროულად, რთულია „დიფფეიკების“ შემქმნელთა და გამავრცელებელთა იდენტიფიცირება და მათი დადგენა. სამართლებრივ გამოწვევას ასევე წარმოადგენს „დიფფეიკის“ ერთიანი, საყოველთაოდ აღიარებული განმარტების არარსებობა. აღნიშნული გამოწვევების დასაძლევად მნიშვნელოვანი დატვირთვა გააჩნია როგორც ეროვნულ კანონმდებლობას, ისე სახელმწიფოთა შორის არსებულ სხვადასხვა საერთაშორისო ინსტრუმენტებს. განსაკუთრებით აღსანიშნავია 2024 წლის ევროკავშირის „ხელოვნური ინტელექტის აქტი“, რომელიც აწესებს გამჭვირვალობის მოთხოვნებს ხელოვნური ინტელექტის მიერ გენერირებული გარკვეული კონტენტისთვის, მათ შორის, „დიფფეიკებისთვის“. თუმცა, უნდა აღინიშნოს, რომ აქტი მოქმედებს ევროკავშირის მასშტაბით. სამართლებრივ მექანიზმებთან ერთად, არანაკლებ მნიშვნელოვანია არასამართებრივი საშუალებებიც. შეიძლება აღინიშნოს, რომ არასამართლებრივი საშუალებები გარკვეულწილად შემავსებელ ხასიათს ატარებენ. მნიშვნელოვანი დატვირთვა გააჩნია ე.წ. „პლატფორმის თვითრეგულირებას“, სხვადასხვა ტექნოლოგიური კომპანიების ჩართულობას, მათ ურთიერთკოორდინაციას, მედიაწიგნიერებას, საზოგადოებაში ხელოვნური

[2] Corresponding author.
E-mail addresses: george.khatidze@ciu.edu.ge (გ. ხატიძე)
ORCID: https://orcid.org/0009-0006-6689-8505

ინტელექტის არსის, მისი რაობის, მის ტექნოლოგიებზე დამყარებული „დიფფეიკების“ შესახებ ცნობიერების ამაღლებას. თუმცა, არასამართლებრივ მექანიზმებს სავალდებულო ძალა არ გააჩნიათ. მთლიანობაში, შეიძლება ითქვას, რომ „დიფფეიკების“ რეგულირების და მოწესრიგების საშუალებები გარკვეულწილად ფრაგმენტირებულია. სახეზეა ერთიანი, უნიფიცირებული მექანიზმების ნაკლებობა.

1. შესავალი

ინტერნეტის გაჩენამ ხელი შეუწყო ხელოვნური ინტელექტისა და შესაბამისი სისტემების არნახული სისწრაფით განვითარებას, რაც დღეს გავლენას ახდენს ჩვენს ყოველდღიურ ცხოვრებაზე. აღნიშნულმა, თავის მხრივ, დააჩქარა ტექნოლოგიური პროცესები და განაპირობა ხელოვნური ინტელექტის სხვადასხვა მოწინავე ტექნოლოგიებზე დამყარებული „დიფფეიკების“ შექმნა და განვითარება. მიუხედავად იმისა, რომ „დიფფეიკების“ გამოყენება შესაძლებელია დადებითი მიზნებისთვის, მაგალითად, კინოინდუსტრიაში, ვიდეოთამაშებში, ფილმების დუბლირებისა და თარგმნისას, სხვადასხვა საგანმანათლებლო, სახელოვნებო და შემეცნებით პროექტებში, მათი შექმნა და გამოყენება მხოლოდ ზემოაღნიშნული მიზნებით არ ხდება. „დიფფეიკები“ საკმაოდ ხშირად გამოიყენება დეზინფორმაციისა და ყალბი ამბების შესაქმნელად და გასავრცელებლად, ფინანსური თაღლითობის მიზნით, პოლიტიკური მანიპულაციისა და პროპაგანდისთვის. „დიფფეიკების“ გამოყენების უარყოფითი პრაქტიკა ასევე უკავშირდება საერთაშორისო საინფორმაციო ომს, რომელიც ითვალისწინებს ინფორმაციით განზრახ მანიპულირებას, კონტროლს და გამოყენებას საზღვრებს მიღმა სტრატეგიული უპირატესობის მოსაპოვებლად. იგი მოიცავს პროპაგანდას, დეზინფორმაციას, კიბეროპერაციებსა და მედია მანიპულირებას. ზემოაღნიშნულიდან გამომდინარე, „დიფფეიკების“ უარყოფითი ეფექტის შემცირებისა თუ აღმოფხვრის მიზნით, იკვეთება მათი სამართლებრივი და არასამართლებრივი მექანიზმებითა და საშუალებებით რეგულირების საჭიროება.

ნაშრომის მიზანს, თავად „დიფფეიკების“ არსისა და მათი უარყოფითი ეფექტების განსაზღვრასთან ერთად, წარმოადგენს რეგულირების როგორც სამართლებრივი, ისე არასამართლებრივი მექანიზმებისა და საშუალებების იდენტიფიცირება, მათი გამოწვევებისა და შესაძლებლობების ანალიზი. რამდენადაც „დიფფეიკები“ ეფუძნება ხელოვნური ინტელექტის სხვადასხვა მოწინავე ტექნოლოგიებს, ნაშრომის პირველ თავში განხილულია ხელოვნური ინტელექტის არსი, მისი განვითარების საფუძვლები, თეორიული და ისტორიული ასპექტები, მისი განმარტება და შემადგენელი ელემენტები,

რომელიც შემოთავაზებულია სხვადასხვა საერთაშორისო ორგანიზაციების მიერ, და ასევე 2024 წელს ევროკავშირში მიღებულ ხელოვნური ინტელექტის აქტში, რომელიც წარმოადგენს მსოფლიოში პირველ ყოვლისმომცველ სამართლებრივ ჩარჩოს და არეგულირებს ხელოვნური ინტელექტის შექმნასა და გამოყენებას. მეორე თავში განხილულია „დიფფეიკების“ რაობა, მისი შექმნის ტექნოლოგიური მექანიზმები, მისი გამოყენების დადებითი, და ასევე უარყოფითი პრაქტიკა, სადაც ყურადღება გამახვილებულია საერთაშორისო საინფორმაციო ომსა და მისი წარმოების სხვადასხვა მეთოდებზე. მესამე თავი ეძღვნება „დიფფეიკების“ სამართლებრივ რეგულირებას, მის საჭიროებას და მასთან მიმართებით არსებულ შესაძლებლობებსა და გამოწვევებს. განხილული და გაანალიზებულია როგორც რიგი ქვეყნების ეროვნული კანონმდებლობა, ისე სხვადასხვა საერთაშორისო ინსტრუმენტები. ნაშრომის მეოთხე თავში განხილულია „დიფფეიკების“ რეგულირების არასამართლებრივი - ტექნოლოგიური და პრაქტიკული საშუალებები. ხაზგასმულია რეგულირების კომპლექსური მიდგომის აუცილებლობა.

2. ლიტერატურის მიმოხილვა

საკვლევ თემატიკასთან მიმართებით, მისი სიახლის, სპეციფიურობისა და ინტერდისციპლინური ხასიათის გათვალისწინებით, არსებული სამეცნიერო ლიტერატურა შედარებით მწირია.

ხელოვნური ინტელექტის არსის გააზრება, მისი განვითარების ისტორიული ასპექტები და ევოლუცია წარმოდგენილია Stuart J. Russell-ის და Peter Norvig-ის მონოგრაფიაში “Artificial Intelligence: A Modern Approach”. ავტორთა მოსაზრებით, ხელოვნური ინტელექტი თავისი არსით ითვალისწინებს ინტერდისციპლინურ ელემენტებს, რომლის განვითარებაში სხვადასხვა მეცნიერებამ თავისი წვლილი შეიტანა. მისი საფუძვლებია: ფილოსოფია, მათემატიკა, ეკონომიკა, ნეირომეცნიერება, ფსიქოლოგია, კომპიუტერული მეცნიერება და ინჟინერია, კიბერნეტიკა, ლინგვისტიკა. ხელოვნური ინტელექტი ასევე ეფუძნება ლოგიკის კანონებს და მიზნად ისახავს სწორი დასკვნების გამოტანას. ხელოვნური ინტელექტის შეფასების კლასიკური მეთოდია ტიურინგის ტესტი, რომელიც მოიცავს ბუნებრივი ენის დამუშავებას, ცოდნის წარმოდგენას, მსჯელობასა და სწავლებას. ზემოაღნიშნული საკითხები ასევე გაანალიზებულია რუმინელ-ამერიკელი ფსიქოლოგის David Wechsler-ის ნაშრომში “The Measurement and Appraisal of Adult Intelligence”; და ინტელექტის კვლევის წამყვანი სპეციალისტებისა და ექსპერტთა ჯგუფის მიერ მომზადებულ სამეცნიერო კვლევაში “Intelligence: Knows and Unknows”. ასევე მნიშვნელოვანია ინგლისელი მათემატიკოსის, კომპიუტერული მეცნიერის, ლოგიკოსის და თეორიული ბიოლოგის ალან ტიურინგის ნაშრომი „გამოთვლითი მაქანები და ინტელექტი“, რომლის მიხედვით ინტელექტი განისაზღვრება ქცევითი კრიტერიუმებით, ფუნქციონალისტური მიდგომით და არა შინაგანი ცნობიერებით. მისი ვარაუდით, მანქანებმა შეიძლება გამოავლინონ

ადამიანებისთვის დამახასიათებელი ქცევა, რაც ადამიანისა და მანქანური ინტელექტის პრაქტიკულ განსხვავებას სულ უფრო გაართულებს.

ნაშრომში გამოყენებული მონოგრაფიების, კვლევებისა და სამეცნიერო სტატიების გარდა, ხელოვნური ინტელექტის სამართლებრივი არსის განსაზღვრისა და ნორმატიული განმარტების კუთხით მნიშვნელოვანია სხვადასხვა საერთაშორისო ორგანიზაციების, მათ შორის, ეკონომიკური თანამშრომლობისა და განვითარების ორგანიზაციისა და იუნესკოს მიერ შემოთავაზებული ფორმულირებები. განსაკუთრებით საყურადღებოა 2024 წელს ევროკავშირში მიღებული ხელოვნური ინტელექტის აქტი. აღნიშნული აქტი მნიშვნელოვანია თავად „დიფფეიკების“ სამართლებრივი რეგულირების თვალსაზრისით, რომელიც მათი გავრცელებისას ითვალისწინებს შესაბამისი სამართლებრივი და ტექნიკური მექანიზმების გამოყენებას.
საერთაშორისო საინფორმაციო ომის არსის, მისი მცირე ისტორიული ექსკურსის, შედარებით ახალი პრაქტიკის შესწავლისა და თავად საერთაშორისო საინფორმაციო ომის წარმოების სხვადასხვა მეთოდების კვლევისა და ანალიზის მიზნით გამოყენებული იქნა როგორც სამეცნიერო ნაშრომები, ისე სხვა წყაროები.

„დიფფეიკების“ ნორმატიული და შედარებით-სამართლებრივი კვლევის კუთხით საინტერესოა Edmarverson A. Santos-ს ნაშრომი „დიფფეიკები საერთაშორისო სამართალში: სამართლებრივი სტატუსი და ხარვეზები“, სადაც ავტორის მოსაზრებით, „დიფფეიკების“ მიერ წარმოშობილი გამოწვევები არა მხოლოდ მათ ტექნიკურ მახასიათებლებსა და ელემენტებში გამოიხატება, არამედ მის გავლენაზე სამართალსა და საერთაშორისო სამართლის ფუნდამენტურ პრინციპებზე. ნაშრომში ასევე გაზგასმულია „დიფფეიკების“ ტრანსსასაზღვრო ბუნებაც, როგორც მათი რეგულირების ერთ-ერთი მნიშვნელოვანი გამოწვევა. ასევე შესწავლილია რიგი ქვეყნების კანონმდებლობა (გაერთიანებული სამეფო, ნიდერლანდები, იაპონია, აშშ, ავსტრალია), პრაქტიკა და საერთაშორისო ინსტრუმენტები - სახელმწიფოთა შორის არსებული ორმხრივი და მრავალმხრივი შეთანხმებებისა თუ კონვენციების სახით, რომელთა მიზანსაც სახელმწიფოთა შორის კიბერდანაშაულის წინააღმდეგ თანამშრომლობა წარმოადგენს. აღნიშნული კუთხით აღსანიშნავია 2024 წელს ევროკავშირში მიღებული ხელოვნური ინტელექტის აქტი, ევროპის საბჭოს კონვენცია „კიბერდანაშაულის შესახებ“ (Convention on Cybercrime), „გაეროს კონვენცია კიბერდანაშაულის წინააღმდეგ“ (United Nations Convention against Cybercrime), „შეთანხმება ამერიკის შეერთებულ შტატებსა და გაერთიანებულ სამეფოს შორის დანაშაულის წინააღმდეგ ბრძოლის მიზნით ელექტრონულ მონაცემებზე წვდომის შესახებ“, “ევროკავშირ-იაპონიის სტრატეგიული პარტნიორობის შეთანხმება”.
„დიფფეიკების“ ტექნოლოგიური ბუნებიდან და შემადგენელი ელემენტებიდან გამომდინარე, შესწავლილია სხვადასხვა კომპანიების - Meta, YouTube, X - რეგულირების

ისეთი პრაქტიკა, როგორიცაა ე.წ. „პლატფორმის თვითრეგულირება“ (Platform Self-Regulation), „YouTube-ის დეზინფორმაციის პოლიტიკა“ (YouTube Misinformation Policies) და „X-ის ავთენტურობის პოლიტიკა“ (X – Authenticity Policy). ასევე ხაზგასასმელია „კონტენტის წარმომავლობისა და ავთენტურობის კოალიციის“ (The Coalition for Content Provenance and Authenticity (C2PA)), როგორც ტექნოლოგიური კომპანიების, მედიაორგანიზაციებისა და მკვლევრების თანამშრომლობის მაგალითი, რომელიც ითვალისწინებს „დიფფეიკების“ გამოვლენის მიზნით სხვადასხვა მონაცემთა ურთიერთგაზიარებას, საერთო ტექნიკური სტანდარტების შემუშავებას, ურთიერთთავსებადი იდენტიფიცირების სტანდარტების შექმნასა და საფრთხეების შესახებ ინფორმაციის ურთიერთგაცვლას.

3. მასალები და მეთოდები

წინამდებარე ნაშრომი შესრულებულია თეორიული და დოკუმენტური კვლევის ფარგლებში. საკვლევი თემა თავისი შინაარსით ინტერდისციპლინურია. გამოყენებულია კვლევის ისტორიული, ნორმატიული და შედარებითი მეთოდები. შესწავლილ იქნა ხელოვნური ინტელექტის განვითარების ისტორია და მისი ევოლუცია, ხელოვნური ინტელექტის სხვადასხვა მოწინავე ტექნოლოგიებზე დამყარებული „დიფფეიკების“ შექმნის ტექნოლოგიური ეტაპები და მათი შემადგენელი ელემენტები. გამოყენებული მონოგრაფიები, სამეცნიერო სტატიები და სხვა კვლევითი ნაშრომები დაცულია შესაბამის ელექტრონულ ბაზებში, რომელიც ხელმისაწვდომია მკითხველთათვის.

კვლევის ფარგლებში გაანალიზებულია სხვადასხვა ქვეყნის კანონმდებლობა, შესაბამისი სამართლებრივი ნორმები და საერთაშორისო სამართლებრივი მექანიზმები. ნორმატიული წყაროები ასახულია ოფიციალურ გამოცემებში, ბაზებსა და ვებგვერდებზე და მკითხველთათვის ღიად არის ხელმისაწვდომი. ასევე განხორციელებულია სხვადასხვა სამართლებრივი და არასამართლებრივი მექანიზმებისა და საშუალებების შედარებითი ანალიზი.

4. შედეგების განხილვა

4.1. ხელოვნური ინტელექტი: არსი და განვითარება

ინტელექტი გულისხმობს ადამიანის აზროვნების უნარს, გონებრივი განვითარებისა და ცოდნის დონეს - იფიქროს და იმოქმედოს რაციონალურად (Lingvisturi Teknologiebis Jgufi, n.d.; Russell & Norvig, 2022). რუმინელ-ამერიკელი ფსიქოლოგის, დავიდ ვექსლერის განმარტებით, „ინტელექტი არის ინდივიდის უნარი, იმოქმედოს მიზანმიმართულად, იაზროვნოს რაციონალურად და ეფექტიანად ადაპტირდეს გარემოსთან“ (Wechsler, 1958). 1996 წელს გამოქვეყნდა ერთ-ერთი მნიშვნელოვანი სამეცნიერო კვლევა - „ინტელექტი: ცნობილი და უცნობი“, რომელიც ინტელექტის კვლევის წამყვანი სპეციალისტებისა და

ავტორიტეტულ ექსპერტთა ჯგუფის მიერ მომზადდა (Neisser et al., 1996). აღნიშნული კვლევის მიხედვით, ინტელექტი წარმოადგენს უნარების ერთობლიობას, რომელიც მოიცავს რთული იდეების გაგებას, გარემოსთან ადაპტაციას, გამოცდილებიდან სწავლას, ლოგიკურ აზროვნებას და პრობლემების გადაჭრას (Neisser et al., 1996). მაშასადამე, ადამიანური ინტელექტის განმარტების ძირითადი კომპონენტებია: მიზანმიმართული მოქმედება, რაც დასახული მიზნების მიღწევის უნარს გულისხმობს; რაციონალური აზროვნება, ანუ ლოგიკური და ანალიტიკური მსჯელობა; გარემოსთან ადაპტაცია, როგორც ახალი და რთული სიტუაციების ეფექტიანად მართვის შესაძლებლობა; ინტელექტი, როგორც მრავალგანზომილებიანი უნარი, კომპლექსური და მრავალფაქტორიანი. ინტელექტი, თავისი არსით წარმოადგენს რთული ბიოლოგიური, ფსიქოლოგიური და სოციალური პროცესების ერთობლიობას (Neisser et al., 1996).

ადამიანური ინტელექტის ასეთი მრავალგანზომილებიანი და რთული ბუნება მნიშვნელოვანი სამეცნიერო ინტერესის საგანია არა მხოლოდ მისი დახასიათების თვალსაზრისით, არამედ მისი ფუნქციონირების იმიტაციის შესაძლებლობის კუთხითაც. სწორედ ამ კონტექსტში ჩნდება კითხვა, შესაძლებელია თუ არა ინტელექტუალური პროცესების მოდელირება ტექნოლოგიური სისტემების საშუალებით, რაც საფუძვლად დაედო ხელოვნური ინტელექტის, როგორც დამოუკიდებელი სამეცნიერო დისციპლინის, ჩამოყალიბებას. აღნიშნული თვალსაზრისით, აქაც, დასაწყისშივე, საინტერესოა განმარტების საკითხი.

უპირველეს ყოვლისა, უნდა აღინიშნოს, რომ დღემდე არ არსებობს ხელოვნური ინტელექტის ერთი, საყოველთაოდ აღიარებული განმარტება. ვხვდებით ხელოვნური ინტელექტის სხვადასხვა განმარტებას, რომელიც შემუშავებულია არაერთი საერთაშორისო ორგანიზაციის მიერ. მათ შორისაა ეკონომიკური თანამშრომლობისა და განვითარების ორგანიზაციის განმარტება, რომლის მიხედვით „ხელოვნური ინტელექტის სისტემა არის მანქანაზე დაფუძნებული სისტემა, რომელიც, ექსპლიციტური ან იმპლიციტური მიზნების მისაღწევად, მიღებული მონაცემების საფუძველზე ადგენს, თუ როგორ შექმნას მისგან მომდინარე ისეთი მონაცემები, როგორიცაა პროგნოზები, რეკომენდაციები ან გადაწყვეტილებები, რომლებსაც შეუძლიათ გავლენა მოახდინონ ფიზიკურ, რეალურად არსებულ ან ვირტუალურ გარემოზე. სხვადასხვა ხელოვნური ინტელექტის სისტემა განსხვავდება ავტონომიისა და ადაპტირების დონით მისი ამოქმედების შემდეგ“ (OECD, 2024). იუნესკოს მიერ შემოთავაზებული განმარტებით კი, „ხელოვნური ინტელექტი გულისხმობს მანქანების უნარს, შეასრულონ მრავალი დავალება, რომლებიც, როგორც წესი, ადამიანის ინტელექტს მოითხოვს, როგორებიცაა: სწავლა, მსჯელობა და პრობლემების გადაჭრა. ეს ტექნოლოგია გავლენას ახდენს სხვადასხვა სფეროზე, მათ შორის, განათლებაზე, სოციალურ მეცნიერებებსა და საინფორმაციო კომუნიკაციაზე“ (UNESCO, 2021).

განსაკუთრებით მნიშვნელოვანია 2024 წელს ევროკავშირში მიღებული ხელოვნური ინტელექტის აქტში არსებული განმარტება, რომელსაც მოგვიანებით განვიხილავთ.

ხელოვნური ინტელექტის ეპოქა 1956 წელს, აშშ-ს შტატ ნიუ-ჰემფშირის ქალაქ ჰანოვერში მდებარე დარტმუთის კოლეჯში გამართულ პატარა, მაგრამ ახლა უკვე კარგად ცნობილი ზაფხულის კონფერენციით დაიწყო (Bringsjord & Govindarajulu, 2018). სწორედ აღნიშნულ კონფერენციაზე დამკვიდრდა ტერმინი „ხელოვნური ინტელექტი“ (“Artificial Intelligence”), რომლის დამაარსებელია ამერიკელი მათემატიკოსი და კომპიუტერული მეცნიერი ჯონ მაკ-კართი (Markoff, 2011).

ხელოვნური ინტელექტის არსში გარკვევის თვალსაზრისით, საკმაოდ მნიშვნელოვანია ინგლისელი მათემატიკოსის, კომპიუტერული მეცნიერის, ლოგიკოსის და თეორიული ბიოლოგის ალან ტიურინგის ნაშრომი „გამოთვლითი მანქანები და ინტელექტი“, რომელიც 1950 წელს გამოქვეყნდა (Turing, 1950). აღნიშნულ ნაშრომში, ალან ტიურინგი გვთავაზობს, რომ შეკითხვა „შეუძლიათ თუ არა მანქანებს ფიქრი?“, რომელიც საკმაოდ ბუნდოვანია, ჩანაცვლდეს ოპერაციული კრიტერიუმით ე.წ. „იმიტაციის თამაშით“ (ტიურინგის ტესტი) (Florkin, 2011), რომლის მიხედვით ინტელექტი განისაზღვრება ქცევითი კრიტერიუმებით, ფუნქციონალისტური მიდგომით და არა შინაგანი ცნობიერებით (Turing, 1950). ტიურინგის ვარაუდით, იმიტაციის თამაშში მანქანებმა შეიძლება გამოავლინონ ადამიანებისთვის დამახასიათებელი ქცევა, რაც ადამიანისა და მანქანური ინტელექტის პრაქტიკულ განსხვავებას სულ უფრო გაართულებს (Zhou, 2025). შეიძლება ითქვას, რომ ალან ტიურინგს ხელოვნური ინტელექტის არსსა და კონცეფციაში გარკვევის თვალსაზრისით ფუნდამენტური წვლილი აქვს შეტანილი.

აქვე, საინტერესოა ამერიკელი ფილოსოფოსის ჯონ ჰოგლენდის ნაშრომები და მოსაზრებები ხელოვნური ინტელექტის შესახებ (Haugeland, 1989). მისი ნაშრომები შეეხება გონების ფილოსოფიას, ხელოვნურ ინტელექტს, კოგნიტურ მეცნიერებას. ჰოგლენდის მოსაზრებით, ადამიანის აზროვნება და კომპიუტერული გამოთვლები შესაძლოა ფუნდამენტურად ერთი და იგივე იყოს (Vellino, 1986). თუ აღნიშნული წარმოადგენს სიმართლეს, მაშინ მანქანას პოტენციურად შეიძლება ჰქონდეს გონება - არა მხოლოდ ინტელექტის იმიტაცია, არამედ სინამდვილეში იყოს ინტელექტუალური (Vellino, 1986). ჰოგლენდის თქმით, ხელოვნური ინტელექტი არის მცდელობა იმის აღმოსაჩენად, არის თუ არა ინტელექტი გამოთვლების ერთგვარი სახეობა და შეიძლება თუ არა თავად ადამიანის გონება იყოს ინფორმაციის დამუშავების სისტემა (Vellino, 1986).
დარტმუთის კონფერენციის შემდეგ ხელოვნური ინტელექტის განვითარებამ სხვადასხვა ეტაპები გაიარა, როგორც სწრაფი პროგრესის, ისე შეფერხების პერიოდები (Russell & Norvig, 2022). მოგვიანებით, 1990-იან წლებში, მანქანური სწავლებისა და გამოთვლითი

შესაძლებლობების ზრდის წყალობით დარგმა განვითარების ახალ ეტაპს მიაღწია (Russell & Norvig, 2022). 2020 წლიდან ხელოვნური ინტელექტი გენერაციული სისტემების ეპოქაში გადავიდა (He et al., 2025).

ხელოვნური ინტელექტი, როგორც კომპიუტერული მეცნიერების დარგი შეისწავლის და ქმნის სისტემებს, რომელთაც შეუძლიათ შეასრულონ ის ამოცანები, რომლის შესასრულებლად საჭიროა ადამიანური ინტელექტი (Russell & Norvig, 2022). ხელოვნური ინტელექტის მიზანია კომპიუტერულ სისტემებში ადამიანის აზროვნების მოდელირება, ის ეფუძნება კოგნიტურ მეცნიერებასა და ფსიქოლოგიას (Russell & Norvig, 2022). ხელოვნური ინტელექტის შეფასების კლასიკური მეთოდია ტიურინგის ტესტი, რომელიც მოიცავს ბუნებრივი ენის დამუშავებას, ცოდნის წარმოდგენას, მსჯელობასა და სწავლებას (Russell & Norvig, 2022). ხელოვნური ინტელექტი ასევე ეფუძნება ლოგიკის კანონებს და მიზნად ისახავს სწორი დასკვნების გამოტანას (Russell & Norvig, 2022).

2024 წელს ევროკავშირში მიღებული იქნა ხელოვნური ინტელექტის აქტი, რომელიც წარმოადგენს მსოფლიოში პირველ ყოვლისმომცველ სამართლებრივ ჩარჩოს და არეგულირებს ხელოვნური ინტელექტის შექმნასა და გამოყენებას (European Parliament & Council of the European Union, 2024). ხელოვნური ინტელექტის აქტი მოქმედებს ევროკავშირის მასშტაბით და მისი მიზნებია უსაფრთხო და სანდო ხელოვნური ინტელექტის უზრუნველყოფა; ადამიანის ფუნდამენტური უფლებების და ღირებულებების დაცვა; ინოვაციისა და ეკონომიკური განვითარების ხელშეწყობა; ევროკავშირში ერთიანი სამართლებრივი ჩარჩოს შექმნა; ეთიკური, გამჭვირვალე და ანგარიშვალდებული ხელოვნური ინტელექტის უზრუნველყოფა (European Commission, n.d.). აქტის მიზანია უზრუნველყოს, რომ ხელოვნური ინტელექტის სისტემები იყოს უსაფრთხო, გამჭვირვალე, ეთიკური და ადამიანის უფლებებთან შესაბამისი, ამასთანავე ხელი შეუწყოს ინოვაციასა და ეკონომიკურ განვითარებას (European Commission, n.d.). ხელოვნური ინტელექტის აქტი წარმოადგენს იმ სამართლებრივ ჩარჩოს, რომელიც აწესებს მოთხოვნებს შესაბამისი დონისა და სახეობის მიხედვით, ხელოვნური ინტელექტის სისტემის გამოყენებასთან დაკავშირებული რისკებისა და გამოწვევების საპასუხოდ (Hunton Andrews Kurth LLP, 2024). აღნიშნულის მიხედვით, რაც უფრო მაღალია ის რისკი, რომელიც ხელოვნური ინტელექტის სისტემის გამოყენებას ახლავს, მით მკაცრია ის სამართლებრივი მოთხოვნები, რომელიც სისტემამ და მისმა გამოყენებამ უნდა დააკმაყოფილოს (Hunton Andrews Kurth LLP, 2024). მიზანი მდგომარეობს ფუნდამენტური უფლებების დაცვაში, უსაფრთხოების უზრუნველყოფასა და ინოვაციების ხელშეწყობაში (Hunton Andrews Kurth LLP, 2024). ამგვარი მიდგომა შეიძლება ითქვას, რომ ერთგვარი ბალანსის ფუნქციას ატარებს ხელოვნური ინტელექტის სისტემების გამოყენების დადებით მხარეებსა და არაკეთილსინდისიერ პრაქტიკას შორის.

აქტი ხელოვნური ინტელექტის სისტემების გამოყენების ოთხ კატეგორიას ადგენს მათი რისკების გათვალისწინებით. პირველი - აკრძალული ხელოვნური ინტელექტის სისტემები, რომელიც მოიცავს, მათ შორის, იმ სისტემებს, რომლებიც იყენებენ ქვეცნობიერ, მიზანმიმართულად მანიპულაციურ ან შეცდომაში შემყვან ტექნიკებს; სისტემები, რომელიც გამოიყენება სოციალური ქულების მინიჭებისთვის, სახის ამოცნობის მონაცემთა ბაზების შესაქმნელად, ფიზიკურ პირთა ემოციების გამოსავლენად სამუშაო ადგილზე ან საგანმანათლებლო დაწესებულებებში, ბიომეტრიული კატეგორიზაციისთვის, რომელიც ეფუძნება მგრძნობიარე მონაცემებს (Bird & Bird, 2024). აღსანიშნავია ისიც, რომ აქტი ითვალისწინებს გამონაკლისებს, რომელიც განსაკუთრებულ შემთხვევებში აკრძალული პრაქტიკის გამოყენების შესაძლებლობას უშვებს. მაგალითად, სამართალდაცვითი მიზნებისთვის, საჯარო ადგილებში რეალურ დროში ბიომეტრიული იდენტიფიკაცია დასაშვებია მხოლოდ გამონაკლის შემთხვევებში და წინასწარი ნებართვის საფუძველზე (Bird & Bird, 2024). მეორე - მაღალი რისკის ხელოვნური ინტელექტის სისტემები, რომელიც ითვალისწინებს ისეთ შემთხვევებს, როდესაც აღნიშნული სისტემების გამოყენებამ შეიძლება მნიშვნელოვნად მავნე ზეგავლენა მოახდინოს ადამიანთა ჯანმრთელობაზე, უსაფრთხოებასა და ფუნდამენტურ უფლებებზე (Bird & Bird, 2024). აღნიშნული კატეგორია თავის მხრივ იყოფა ორ ქვეკატეგორიად. მესამე - გამჭვირვალობის მოთხოვნებს დაქვემდებარებული ხელოვნური ინტელექტის სისტემები, რომელიც წარმოადგენს ისეთ სისტემებს, რომელთაც ახასიათებთ გამჭვირვალობასთან დაკავშირებული რისკები ან შეუძლიათ მომხმარებელთა შეცდომაში შეყვანა (Hunton Andrews Kurth LLP, 2024). ასეთ შემთხვევებში, ხელოვნური ინტელექტის აქტი ავალდებულებს ხელოვნური ინტელექტის სისტემების მიმწოდებლებს, აღნიშნული რისკების შემცირების მიზნით დაიცვან სპეციალური წესები (Bird & Bird, 2024). მაგალითად, მომხმარებლების ინფორმირება ხელოვნური ინტელექტის გამოყენების შესახებ, მარკირების ვალდებულება (Hunton Andrews Kurth LLP, 2024). აღნიშნული კატეგორია მოიცავს ხელოვნური ინტელექტის სისტემებს, რომლებიც განკუთვნილია ინდივიდებთან პირდაპირი ინტერაქციისთვის; ხელოვნური ინტელექტის სისტემები, მათ შორის, ზოგადი დანიშნულების ხელოვნური ინტელექტის სისტემები, რომლებიც ქმნიან სინთეზურ აუდიოს, გამოსახულებას, ვიდეოს ან ტექსტურ კონტენტს; ემოციების ამოცნობისა ან ბიომეტრიული კატეგორიზაციის სისტემები; ხელოვნური ინტელექტის სისტემები, რომლებიც ქმნიან ან მანიპულირებენ გამოსახულებას, აუდიო ან ვიდეოკონტენტს და წარმოადგენენ ე.წ. „დიფფეიკებს“ (ModelOp, n.d.). მეოთხე - ზოგადი დანიშნულების ხელოვნური ინტელექტის მოდელები. ხელოვნური ინტელექტის მოდელები წარმოადგენენ ხელოვნური ინტელექტის სისტემების აუცილებელ კომპონენტებს, თუმცა განცალკევებით, ისინი თავად არ არიან ხელოვნური ინტელექტის სისტემები (Hunton Andrews Kurth LLP, 2024). იმისათვის, რომ ხელოვნური ინტელექტის მოდელები ჩამოყალიბდნენ ხელოვნური ინტელექტის სისტემებად, მოდელებს სჭირდებათ

სხვა კომპონენტების დამატება, როგორიცაა, მაგალითად, მომხმარებლის ინტერფეისი (Hunton Andrews Kurth LLP, 2024). ხელოვნური ინტელექტის აქტი არეგულირებს ხელოვნური ინტელექტის სისტემებს და არა მოდელებს, თუმცა იგი ითვალისწინებს გარკვეულ წესებს ზოგადი დანიშნულების ხელოვნური ინტელექტის მოდელებთან დაკავშირებით (ModelOp, n.d.).

საკმაოდ საინტერესოა აღნიშნული აქტი ხელოვნური ინტელექტის განმარტების თვალსაზრისით. აქტით შემოთავაზებულია ხელოვნური ინტელექტის სისტემის (AI system) განმარტება, რომლის მიხედვით „ხელოვნური ინტელექტის სისტემა„ ნიშნავს მანქანაზე დაფუძნებულ სისტემას, რომელიც შექმნილია ავტონომიის სხვადასხვა დონით ფუნქციონირებისთვის და შესაძლოა გამოავლინოს ადაპტირების უნარი დანერგვის შემდეგ, და რომელიც, აშკარა ან ნაგულისხმევი მიზნების მისაღწევად, მიღებული მონაცემების საფუძველზე განსაზღვრავს, თუ როგორ უნდა შექმნას ისეთი შედეგები, როგორებიცაა: პროგნოზები, შინაარსი, რეკომენდაციები ან გადაწყვეტილებები, რომლებიც გავლენას ახდენენ რეალურად არსებულ ან ვირტუალურ გარემოზე“ (European Parliament & Council of the European Union, 2024). აღნიშნული განმარტების არაერთი შეფასება, მათ შორის, კრიტიკა არსებობს. ხელოვნური ინტელექტის სისტემა განმარტებულია როგორც „...მანქანაზე დაფუძნებული სისტემა...“. ერთ-ერთი მოსაზრებით, აღნიშნული განმარტებით რთულია ერთმანეთისგან გაიმიჯნოს ხელოვნური ინტელექტის (AI) სისტემები არახელოვნური ინტელექტის პროგრამული უზრუნველყოფებისგან (Nessler & Wendehorst, 2024). აღნიშნული, თავის მხრივ, სამართლებრივ ბუნდოვანებას წარმოშობს და გადაჭარბებული რეგულირებისა და არათანმიმდევრული გამოყენების რისკს ქმნის (Muller & Teilhard De Chardin, 2025). რიგი შეფასებებით, განმარტება საკმაოდ ფართოა, მასში გამოყენებულ სხვადასხვა ელემენტებს არ გააჩნიათ ტექნიკური საფუძვლები და მნიშვნელობა (Linklaters, 2025). ერთ-ერთი კვლევის მიხედვით, აქტში ხელოვნური ინტელექტის მოცემული განმარტება ფართო და, გარკვეულწილად, ბუნდოვანია, შესაბამისად, საჭიროებს დამატებით განმარტებას იმისათვის, რომ უზრუნველყოფილი იქნეს სამართლებრივი სიცხადე, თავიდან იქნეს აცილებული ზედმეტი რეგულაცია და მოხდეს მისი ერთგვაროვანი გამოყენება (Irish Council for Civil Liberties, 2025).

მიუხედავად განმარტებასთან მიმართებით არსებული სხვადასხვა მოსაზრებისა, ხელოვნური ინტელექტის აქტი მნიშვნელოვანი დოკუმენტია, სამართლებრივი და არა მხოლოდ აღნიშნული თვალსაზრისით. ხელოვნური ინტელექტის განვითარების ტენდენციებმა და არსებულმა დონემ ცხადყო მისი რეგულირების საჭიროება. მართალია აქტი არ წარმოადგენს საყოველთაო საერთაშორისო სამართლებრივ ინსტრუმენტს და ის ევროკავშირის მასშტაბით მოქმედებს, თუმცა მისი შემუშავება და მიღება უმნიშვნელოვანესი ნაბიჯია იმ სამართლებრივი, ინსტიტუციური, ტექნიკური თუ სხვა

მექანიზმებისა და საშუალებების ჩამოყალიბებისა და შემდგომი განვითარების თვალსაზრისით, რომელთაც ერთგვარი ბალანსის ფუნქცია გააჩნიათ ხელოვნური ინტელექტის სისტემების გამოყენების დადებით მხარეებსა და არაკეთილსინდისიერ პრაქტიკას შორის, მათ შორის ფუნდამენტური უფლებების დარღვევების, დეზინფორმაციის, ე.წ. „დიფფეიკების“ და სხვა ქმედებათა წინააღმდეგ.

ხელოვნური ინტელექტი, თავისი არსით, ითვალისწინებს ინტერდისციპლინურ ელემენტებს, რომლის განვითარებაში სხვადასხვა მეცნიერებამ თავისი წვლილი შეიტანა (Russell & Norvig, 2022). მისი საფუძვლებია ფილოსოფია, მათემატიკა, ეკონომიკა, ნეირომეცნიერება, ფსიქოლოგია, კომპიუტერული მეცნიერება და ინჟინერია, კიბერნეტიკა, ლინგვისტიკა (Russell & Norvig, 2022).

4.2. „დიფფეიკები“

4.2.1. რა არის „დიფფეიკები“ და როგორ იქმნება ის?

„დიფფეიკები“ არის სინთეზური მედია, მათ შორის, ფოტოები, ვიდეოები ან აუდიოკონტენტი, რომელიც გენერირებულია ხელოვნური ინტელექტის სისტემებით და ასახავს იმას, რაც სინამდვილეში არ არსებობს ან მოვლენებს, რომელიც არ მომხდარა (Payne, 2026). „დიფფეიკებით“ ხდება გარკვეული მოვლენების იმიტაცია. ტერმინი „დიფფეიკი“ აერთიანებს ორ ძირითად ელემენტს - „დიფ“ “deep” (ღრმა), რომელიც გამომდინარეობს ხელოვნური ინტელექტის ღრმა სწავლების ტექნოლოგიიდან და, თავის მხრივ, წარმოადგენს მანქანური სწავლების სახეობას, რომელიც ითვალისწინებს დამუშავების სხვადასხვა დონეს, და „ფეიკ“ “fake” (ყალბი, არარეალური), რაც გულისხმობს იმას, რომ კონტენტი არარეალურია (Payne, 2026). აღნიშნული ტერმინი პირველად გამოყენებული იქნა ერთ-ერთი პოპულარული სოციალური ინტერნეტპლატფორმის რედდითის (Reddit) მიერ სინთეზური მედიის აღსანიშნავად, როდესაც პლატფორმის მოდერატორმა შექმნა ქვერედიტი სახელწოდებით “r/deepfakes” და დაიწყო ვიდეოების გამოქვეყნება, სადაც გამოყენებული იყო სახის შეცვლის ტექნოლოგია (Nagumotu, 2022).

„დიფფეიკები“ ეფუძნება ხელოვნური ინტელექტის მოწინავე ტექნოლოგიებს, როგორებიცაა: ღრმა სწავლება (Deep Learning), გენერაციული ხელოვნური ინტელექტი (Generative AI) (Mitra et al., 2024), გენერაციული კონკურენტული ქსელები (GAN’s - Generative Adversarial Networks) (Yasar et al., 2025) და დიფუზორული მოდელები (Diffusion Models) (Britannica Editors, 2026). „დიფფეიკების“ შექმნა რამდენიმე ეტაპს მოიცავს. პირველ რიგში, ხდება მონაცემების შეგროვება, როგორებიცაა: ფოტოები, ვიდეოები და ხმოვანი ჩანაწერები (Yasar et al., 2025). მეორე ეტაპია ხელოვნური ინტელექტის მოდელების ე.წ. „გაწვრთნა“ ისეთი ინფორმაციის შესასწავლად, როგორებიცაა: სახის ნაკვთები, გამომეტყველება და ხმის ნიმუშები (Yasar et al., 2025). შემდეგი ეტაპია სინთეზური მედიის

გენერირება, რომელიც ითვალისწინებს სახის შეცვლას, ხმის კლონირებასა და ტუჩების სინქრონიზაციის მანიპულირებას (Yasar et al., 2025). ბოლო ეტაპია ე.წ. „დახვეწა“, „სრულყოფა“, რათა მაქსიმალურად ზუსტად მოხდეს „დიფფეიკ“ კონტენტის რეალობასთან დაახლოება (Yasar et al., 2025). „დიფფეიკები“ საკმაოდ მძლავრი ინსტრუმენტია, რადგან ისინი ქმნიან რეალობის აღქმის იმიტაციას და აღნიშნულის შედეგად თანამედროვე „დიფფეიკები“ ძნელად გასარჩევია რეალური კონტენტისგან (Croitoru et al., 2024). ისინი იქმნება ავტომატიზაციის პრინციპით. ის, რისი შექმნაც ადრე მოითხოვდა ხანგრძლივ დროს, დღეს უკვე ხელოვნური ინტელექტის თანამედროვე სისტემებით იქმნება რამდენიმე წუთში (Popa et al., 2025). ასევე, ხაზგასასმელია მათი შექმნის მასშტაბურობაც, რაც უმცირეს დროში ათასობით ყალბი კონტენტის შექმნას ითვალისწინებს (Croitoru et al., 2024).

ხელოვნური ინტელექტის ტექნოლოგიებზე დაფუძნებული „დიფფეიკების“ გამოყენება სხვადასხვა მიზნით არის შესაძლებელი. მაგალითად, კინოინდუსტრიაში, ვიდეოთამაშებში, ფილმების დუბლირებისა და თარგმნისას, სხვადასხვა საგანმანათლებლო მიმართულებისას, ისტორიულ რეკონსტრუქციებში, მედიცინაში (O’Brien, 2025). ეს და სხვა მსგავსი შემთხვევები მათ პოზიტიურ გამოყენებას გულისხმობს. თუმცა, „დიფფეიკების“, როგორც მძლავრი ინსტრუმენტების, გამოყენება შესაძლებელია არაკეთილსინდისიერი მიზნებით. აღნიშნული „დიფფეიკების“ გამოყენების უარყოფით პრაქტიკას გულისხმობს, რომელიც დღეს საკმაოდ მასშტაბურია.

4.2.2. „დიფფეიკები“ საერთაშორისო საინფორმაციო ომში

„დიფფეიკები“ საკმაოდ ხშირად გამოიყენება დეზინფორმაციისა და ყალბი ამბების შესაქმნელად და გასავრცელებლად, ფინანსური თაღლითობის მიზნით, პოლიტიკური მანიპულაციისა და პროპაგანდისთვის და სხვა შემთხვევებისა და მიზნებისთვის (Yasar et al., 2025). პოლიტიკაში „დიფფეიკები“ შეიძლება გამოყენებული იქნეს საჯარო პირთა ყალბი ვიდეოების შესაქმნელად, რაც, თავის მხრივ, სხვა ფაქტორებთან ერთად, იწვევს დეზინფორმაციის გავრცელებას, ასუსტებს ნდობას სახელმწიფო ინსტიტუტების მიმართ და შესაძლოა ძალადობისა ან კონფლიქტის პროვოცირებაც მოახდინოს (Vinod & Subramanian, 2024). „დიფფეიკები“ შესაძლოა გამოყენებული იქნეს არჩევნებზე, საზოგადოებრივ აზრზე, პოლიტიკურ პროცესებსა და გადაწყვეტილებებზე გავლენის მოსახდენად (Easttom, 2025). რამდენადაც ხელოვნური ინტელექტის ტექნოლოგიები სწრაფად ვითარდება, მით მეტად იხვეწება „დიფფეიკები“, რაც, თავის მხრივ, მნიშვნელოვან გამოწვევას წარმოადგენს დემოკრატიისთვის (Pawelec, 2022), რადგან, როგორც აღვნიშნეთ, შესწევთ უნარი გამოიწვიონ სახელმწიფო ინსტიტუტების მიმართ საზოგადოების ნდობის შესუსტება (KPMG, n.d.), საფრთხე შეუქმნან ინდივიდებსა და მთლიანად საზოგადოებას, უარყოფითი გავლენა იქონიოს საზოგადოებრივ უსაფრთხოებაზე, სახელმწიფოთა შორის არსებულ

დიპლომატიურ ურთიერთობებზე, ეროვნულ უსაფრთხოებაზე და ჟურნალისტურ საქმიანობაზე (Chesney & Citron, 2019; Diel et al., 2025).

„დიფფეიკების“ გამოყენების უარყოფითი პრაქტიკა ასევე უკავშირდება საერთაშორისო საინფორმაციო ომს, რომელიც ითვალისწინებს ინფორმაციით განზრახ მანიპულირებას, კონტროლს და გამოყენებას საზღვრებს მიღმა სტრატეგიული უპირატესობის მოსაპოვებლად (Whitaker, 2003). იგი მოიცავს პროპაგანდას, დეზინფორმაციას, კიბეროპერაციებსა და მედიამანიპულირებას (LegalClarity Team, 2026). საერთაშორისო საინფორმაციო ომი გულისხმობს როგორც სახელმწიფო, ისე არასახელმწიფო აქტორების მიერ ინფორმაციის, საკომუნიკაციო სისტემების, ნარატივების, მედიისა და ციფრული ტექნოლოგიების გამოყენებას პოლიტიკურ გადაწყვეტილებებსა და შედეგებზე, საზოგადოებრივ აღქმებზე, საზოგადოებრივი აზრის ფორმირებაზე გავლენის მოსახდენად (NATO, 2025). ასეთ შემთხვევაში, ტრადიციული ომისგან განსხვავებით, ძირითად სამიზნეს ფიზიკური ინფრასტრუქტურის ნაცვლად ადამიანის ცნობიერება, საზოგადოებრივი აზრი, ნდობა და გადაწყვეტილებების მიღების პროცესი წარმოადგენს (NATO, 2025).

საერთაშორისო საინფორმაციო ომი არ წარმოადგენს ე.წ. „ახალ ფენომენს“. მისი ისტორიული მაგალითები, ფსიქოლოგიური ოპერაციების სახით, არსებობს მეორე მსოფლიო ომის პერიოდში, ასევე პროპაგანდის წარმოება ცივი ომის დროს (Norwich University, n.d.). ასევე აღსანიშნავია სოციალური მედიის პლატფორმების გამოყენება დეზინფორმაციის გავრცელებისა და საზოგადოებრივ აზრზე გავლენის მოხდენის მიზნით რუსეთ-უკრაინის ომის შემთხვევაში (Premakanthan, 2024). ერთ-ერთი მაგალითია 2022 წლის 16 მარტს, უკრაინაში რუსეთის შეჭრიდან სამი კვირის შემდეგ, „გატეხილ“ უკრაინულ საინფორმაციო ვებ-გვერდზე გავრცელებული ვიდეო „დიფფეიკის“ სახით, რომელიც გავრცელდა Facebook-სა და რუსულ სოციალურ ქსელებში (Detect Deepfakes, 2026). ვიდეოში, სავარაუდოდ, ნაჩვენები იყო პრეზიდენტი ზელენსკი, რომელიც მოუწოდებდა უკრაინელ ჯარისკაცებს იარაღის დაყრისა და დანებებისკენ (Detect Deepfakes, 2026). დღის ბოლოს ის უარყოფილი იქნა უკრაინელი ჩინოვნიკების, თავად ზელენსკისა და ფაქტების შემოწმების რამდენიმე ორგანიზაციის მიერ (Detect Deepfakes, 2026). აღსანიშნავია ასევე ე.წ. „Wolf News“ ფიქციური მედიასაშუალება, რომელიც გამოიყენებოდა სახელმწიფოსთან დაკავშირებულ დეზინფორმაციულ კამპანიაში ჩინეთის კომუნისტური პარტიის პროპაგანდის გასავრცელებლად (Agence France-Presse, 2023). მკვლევრების მიერ იდენტიფიცირებული იქნა ხელოვნური ინტელექტის მიერ გენერირებული „წამყვანები“, რომელიც შექმნილი იყო „დიფფეიკ“ ტექნოლოგიის გამოყენებით, რათა გაეკეთებინათ პოლიტიკური კომენტარი, რომელიც პოპულარიზაციას უწევდა ჩინეთს და აკრიტიკებდა აშშ-ს (Agence France-Presse, 2023).

მკვლევართა მოსაზრებით, საინფორმაციო ომი მოიცავს სამ ძირითად მეთოდს, კერძოდ, პირველი - ფსიქოლოგიურ ოპერაციებს, რომელიც გავლენას ახდენს დამოკიდებულებებზე გარკვეული საკითხებისა და მოვლენების მიმართ, შეხედულებებსა და ქცევებზე; მეორე - ციფრული ომს, რომელიც ითვალისწინებს მოწინააღმდეგის სხვადასხვა სისტემებში, მათ შორის, საკომუნიკაციო სისტემებში ჩარევას; და მესამე - კიბერომს, რომელიც გულისხმობს კომპიუტერულ სისტემებზე, ქსელებსა და შესაბამის ინფრასტრუქტურაზე შეტევას (Premakanthan, 2024). ტექნოლოგიების განვითარების კვალდაკვალ ცხადი ხდება, რომ ხელოვნური ინტელექტი, მანქანური სწავლება, გენერაციული ხელოვნური ინტელექტი და სხვა ტექნოლოგიური საშუალებები გაცილებით გააძლიერებს საინფორმაციო ომის წარმოებას (Premakanthan, 2024). სწორედ ამ თვალსაზრისით, ხელოვნური ინტელექტის სისტემები, რომელიც „დიფფეიკების“ შექმნის ტექნოლოგიურ საშუალებას წარმოადგენს, მნიშვნელოვანი იარაღია საერთაშორისო საინფორმაციო ომის საწარმოებლად (Marcellino et al., 2023). აღნიშნული გამოწვევების საპასუხოდ მნიშვნელოვანია კიბერუსაფრთხოების გაძლიერება, დეზინფორმაციისგან დაცვის გაუმჯობესება და შესაბამისი სამართლებრივი, მათ შორის, საერთაშორისო ინსტრუმენტების განვითარება (Premakanthan, 2024).

4.3. „დიფფეიკების“ სამართლებრივი რეგულირება და მისი გამოწვევები

„დიფფეიკების“ გამოყენება, როგორც უკვე აღვნიშნეთ, შესაძლებელია დადებითი თვალსაზრისით. მაგალითად, მისი გამოყენება კინოინდუსტრიაში, ვიდეოთამაშებში, ფილმების დუბლირებისა და თარგმნისას, სხვადასხვა საგანმანათლებლო, სახელოვნებო და შემეცნებით პროექტებში (Chesney & Citron, 2019; O'Brien, 2025). თუმცა, „დიფფეიკების“ შექმნა და მათი გამოყენება მხოლოდ ზემოაღნიშნული მიზნებით არ ხდება. როგორც აღინიშნა, „დიფფეიკები“ მძლავრ იარაღს წარმოადგენს საერთაშორისო საინფორმაციო ომის საწარმოებლად.

ხელოვნური ინტელექტის და მასზე დაფუძნებული სხვადასხვა სისტემების განვითარებამ, უამრავ დადებით შედეგთან ერთად, საზოგადოებისთვის არაერთი გამოწვევა შექმნა. ხელოვნურმა ინტელექტმა თავისი გავლენა მოახდინა საზოგადოებაზე, ჩვენს ყოველდღიურ ცხოვრებასა და მთლიანად საზოგადოებრივ ურთიერთობებზე. შესაბამისად, ხელოვნური ინტელექტის მასშტაბების, შესაძლებლობების და მისი განვითარების მაჩვენებლების გათვალისწინებით ცხადი ხდება, რომ საჭიროა მისი რეგულირება როგორც სამართლებრივი საშუალებებით, ისე ეთიკური თვალსაზრისით და სხვა ეფექტიანი მექანიზმების გამოყენებით.

სამართლებრივი რეგულირების ჭრილში, რიგი მკვლევრების მოსაზრებით, „დიფფეიკების“ მიერ წარმოშობილი გამოწვევები არა მხოლოდ მათ ტექნიკურ მახასიათებლებსა და ელემენტებში გამოიხატება, არამედ მის გავლენაზე სამართალსა და საერთაშორისო

სამართლის ფუნდამენტურ პრინციპებზე (Santos, 2025). გამოწვევას წარმოადგენს „დიფფეიკების“ ტრანსსასაზღვრო ბუნებაც (Santos, 2025). „დიფფეიკები“ იქმნება ერთ იურისდიქციაში, მისი გავრცელება ხდება გლობალური პლატფორმების მეშვეობით და შესაძლოა მოიცავდეს ერთდროულად რამდენიმე სახელმწიფოს (Santos, 2025). სახელმწიფოთა შორის არსებობს სხვადასხვა საერთაშორისო ინსტრუმენტი, ორმხრივი და მრავალმხრივი შეთანხმებების თუ კონვენციების სახით, რომელთა მიზანსაც სახელმწიფოთა შორის კიბერდანაშაულის წინააღმდეგ თანამშრომლობა წარმოადგენს.

აღსანიშნავია ევროპის საბჭოს კონვენცია „კიბერდანაშაულის შესახებ“ (Convention on Cybercrime) (Council of Europe, 2001). კონვენცია მიღებული იქნა 2001 წელს და წარმოადგენს კიბერდანაშაულის წინააღმდეგ ბრძოლის პირველ კონვენციას (Council of Europe, 2001). კონვენცია განსაზღვრავს მონაწილე სახელმწიფოების მოქმედებებს ციფრული მტკიცებულებების შეგროვების და ტრანსსასაზღვრო ურთიერთთანამშრომლობის კუთხით (Jejelola, 2024). ასევე „გაეროს კონვენცია კიბერდანაშაულის წინააღმდეგ“ (United Nations Convention against Cybercrime) (United Nations Office on Drugs and Crime, n.d.). კონვენცია მიღებული იქნა 2024 წელს და მისი მიზანია საერთაშორისო თანამშრომლობის გაძლიერება ინფორმაციული და საკომუნიკაციო ტექნოლოგიების სისტემების საშუალებით ჩადენილი გარკვეული დანაშაულების წინააღმდეგ საბრძოლველად (United Nations Office on Drugs and Crime, n.d.). აღსანიშნავია სახელმწიფოთა შორის ორმხრივი თანამშრომლობაც, რომელიც ეფუძნება „ორმხრივ შეთანხმებებს სამართლებრივი ურთიერთდახმარების შესახებ“. მაგალითად, „შეთანხმება ამერიკის შეერთებულ შტატებსა და გაერთიანებულ სამეფოს შორის დანაშაულის წინააღმდეგ ბრძოლის მიზნით ელექტრონულ მონაცემებზე წვდომის შესახებ“ (Foreign, Commonwealth & Development Office, 2024), „ევროკავშირ-იაპონიის სტრატეგიული პარტნიორობის შეთანხმება“, რომელიც სხვა საკითხებთან ერთად ითვალისწინებს თანამშრომლობას უსაფრთხოების, ტერორიზმის და კიბერუსაფრთხოების კუთხით (European Union & Japan, 2018).

ერთ-ერთი სირთულე „დიფფეიკების“ სამართლებრივ რეგულირებასთან მიმართებით არის ის, რომ „დიფფეიკები“ არ არის კრიმინალიზებული. ცხადია, რომ მისი დადებითი მიზნებით გამოყენებიდან გამომდინარე, ეს ვერ მოხდება. „დიფფეიკები“ აღქმულია ტექნოლოგიის ჭრილში. მათი გამოყენებით ჩადენილი დანაშაულები და სხვა ქმედებები, როგორც წესი, უფრო ფართო კუთხით განიხილება, მაგალითად კიბერდანაშაული, თაღლითობა, არჩევნებში ჩარევა და სხვა. საერთაშორისო შეთანხმებების უმეტესობაში „დიფფეიკები“ პირდაპირ არ არის ნახსენები. შეიძლება ითქვას, რომ „დიფფეიკების“ რეგულირება არაპირდაპირი ფორმით ხდება, რამდენადაც საერთაშორისო ინსტრუმენტები ქმნიან სამართლებრივ ჩარჩოს, რომლის მეშვეობითაც ქვეყნები თანამშრომლობენ, როდესაც „დიფფეიკები“ გამოიყენება ისეთი ქმედებების განსახორციელებლად, როგორიცაა

თაღლითობა, გამოძალვა და სხვა. დღეს, სამართლებრივი მექანიზმები „დიფფეიკების“ რეგულირების და მართვის თვალსაზრისით განვითარების საკმაოდ ადრეულ ეტაპზეა (Shirish & Komal, 2024). რიგი ქვეყნების ეროვნული კანონმდებლობით განსაზღვრულია სინთეზური მედიის მარკირების ვალდებულება, ასევე დანაშაულად მიიჩნევა განსაკუთრებით მავნე „დიფფეიკების“ შექმნა.

როდესაც სამართლებრივი რეგულირების ჭრილში ვსაუბრობთ „დიფფეიკების“ შექმნასა და გავრცელებასთან დაკავშირებით სამართლებრივ პასუხისმგებლობაზე, ერთ-ერთი სირთულე ასევე უკავშირდება „დიფფეიკების“ შემქმნელთა და გამავრცელებელთა იდენტიფიცირებასა და მათ დადგენას. „დიფფეიკების“ შექმნა და მათი გავრცელება ხდება შესაბამისი ტექნოლოგიების გამოყენებით, ინტერნეტსივრცეში. სწორედ ამიტომ, ტექნოლოგიური პროცესები და საშუალებები იძლევა ასეთ პირთა ანონიმურად დარჩენის შესაძლებლობას, მათი შემდგომი იდენტიფიცირებისა და დადგენის გარეშე (Chesney & Citron, 2019).

ზოგიერთმა ქვეყნებმა შეიმუშავეს შესაბამისი კანონმდებლობა, რომელიც პირდაპირ თუ ირიბად არეგულირებს „დიფფეიკებით“ გამოწვეულ ისეთ შედეგებს, როგორიცაა იმიტაცია, არჩევნების მანიპულირება, დეზინფორმაცია, კონფიდენციალურობის დარღვევა და სხვა (Detect Deepfakes, n.d.). აღნიშნულ აქტებს შორისაა გაერთიანებული სამეფოს “Online Safety Act”, რომელიც მიღებული იქნა 2023 წელს (Department for Science, Innovation and Technology, 2025), ავსტრალიაში მიღებული „სისხლის სამართლის კოდექსის შესწორების შესახებ 2024 წლის კანონი“ (Parliament of Australia, 2024), ნიდერლანდები, სადაც „დიფფეიკების“ სამართლებრივი რეგულირების საფუძველია 2024 წელს ევროკავშირში მიღებული „ხელოვნური ინტელექტის აქტი“, ნიდერლანდების „სისხლის სამართლის კოდექსი“ და „მონაცემთა დაცვის ზოგადი რეგულაცია“ (General Data Protection Regulation) (Detect Deepfakes, 2026), იაპონია, სადაც მოქმედებს კანონი „პერსონალური ინფორმაციის დაცვის შესახებ“, კანონი „ცილისწამების შესახებ“ და 2024 წლის სახელმძღვანელო პრინციპები „ხელოვნური ინტელექტის მიერ გენერირებული კონტენტის შესახებ“ (Detect Deepfakes, 2026).

აქვე, აღსანიშნავია ამერიკის შეერთებული შტატების პრაქტიკა „დიფფეიკების“ საკანონმდებლო რეგულირებასთან დაკავშირებით. 2023 წელს, აშშ-ს კონგრესის წარმომადგენელთა პალატაში ინიცირებული იქნა „დიფფეიკების ანგარიშვალდებულების აქტი“ (Deepfakes Accountability Act), რომლის მიზანიც გახლდათ სამართლებრივი პასუხისმგებლობის დადგენა მავნე „დიფფეიკების“, და ადამიანისა და ეროვნული უსაფრთხოების დაცვის უზრუნველყოფა ხელოვნური ინტელექტის მიერ გენერირებული ყალბი მედიისგან (United States Congress, 2023). აქტის პროექტში გათვალისწინებული იქნა

რამდენიმე მნიშვნელოვანი საკითხი, რომელთა შორისაა: ა) „დიფფეიკების“ გამჭვირვალობა, რომელიც გულისხმობს მათი შემქმნელების მიერ „დიფფეიკების“ მარკირებას, როგორც ერთგვარ აღნიშვნას, რომ კონტენტი წარმოადგენს „დიფფეიკს“; ბ) მავნე „დიფფეიკებთან“ მიმართებით შესაბამისი ზომების განსაზღვრა, რომელიც გულისხმობს სისხლისსამართლებრივი პასუხისმგებლობის დაწესებას არამარკირებული „დიფფეიკების“ შექმნის ან გავრცელებისთვის, რამაც გამოიწვია შევიწროება, ძალადობა, არჩევნებსა და სამთავრობო უწყებების საქმიანობაში ჩარევა, თაღლითობა, უცხოური გავლენის გავრცელება; გ) დაზარალებულთათვის სამართლებრივი მისაგებლის (legal remedy) ხელმისაწვდომობა (United States Congress, 2023). აქტი ამავდროულად ითვალისწინებდა „დიფფეიკების“ განმარტებას, რომლის თანახმად „დიფფეიკი“ ნიშნავს „ნებისმიერ ვიდეო ჩანაწერს, კინოფილმს, ციფრულ გამოსახულებას, ან ფოტოსურათს, ან საუბრის/ქცევის შექმნას და გადმოცემას ტექნოლოგიური საშუალებებით, რომელიც ასახავს იმ პირთა საუბარს ან ქცევას, რომელთაც სინამდვილეში არ უსაუბრიათ და არ უმოქმედიათ იმგვარად, როგორც ეს გადმოცემულია; და რომლის შექმნა არსებითად დამოკიდებული იყო ტექნოლოგიურ საშუალებებზე, და არა სხვა პირის ფიზიკური ან ვერბალური იმიტაციის უნარზე“ (United States Congress, 2023). „დიფფეიკების ანგარიშვალდებულების აქტი“ არ იქნა მიღებული.

ასევე აღსანიშნავია, რომ 2022 წელს, პრეზიდენტმა ბაიდენმა ჩამოაყალიბა სპეციალური სამუშაო ჯგუფი კიბერდანაშაულის წინააღმდეგ ბრძოლის მიზნით, რომლის მისიაც გახლდათ ონლაინ შევიწროებისა და ძალადობის პრევენცია, მათზე რეაგირება და მათთან გამკლავება (White House Task Force to Address Online Harassment and Abuse, 2024). სამუშაო ჯგუფის შექმნა ეფუძნება იმ რეალობას, რომ ონლაინ ძალადობა არ არის მხოლოდ ინტერნეტ სივრცეში არსებული პრობლემა, არამედ წარმოადგენს გამოწვევას საზოგადოებრივი უსაფრთხოებისა და ჯანმრთელობისთვის, სამოქალაქო უფლებებისა და დემოკრატიული მმართველობისთვის (White House Task Force to Address Online Harassment and Abuse, 2024). აღნიშნული სამუშაო ჯგუფის რეგულირების სფეროსგან განსხვავებით, „დიფფეიკების ანგარიშვალდებულების აქტი“ (Deepfakes Accountability Act), რომელიც არ იქნა მიღებული, ითვალისწინებდა გამჭვირვალობისა და პასუხისმგებლობის გაცილებით მძლავრ და ფართო სამართლებრივ მექანიზმებს „დიფფეიკებთან“ მიმართებით, ხოლო სამუშაო ჯგუფმა ორიენტაცია მოახდინა „დიფფეიკებზე“, როგორც ტექნოლოგიით ხელშეწყობილი გენდერული ძალადობის ფორმაზე, და ყოვლისმომცველი „დიფფეიკ“ რეგულაციის შექმნის ნაცვლად ფოკუსირდა პრევენციაზე, მსხვერპლთა მხარდაჭერაზე, სამართალდაცვასა და ხელოვნური ინტელექტის პასუხისმგებლიან მმართველობაზე (United States Congress, 2023). „დიფფეიკების“ სამართლებრივ რეგულირებასთან მიმართებით, ცალკე აღსანიშნავია 2024 წლის ევროკავშირის „ხელოვნური ინტელექტის აქტი“, რომელიც მოქმედებს ევროკავშირის მასშტაბით და როგორც უკვე აღვნიშნეთ აწესებს გამჭვირვალობის მოთხოვნებს ხელოვნური

ინტელექტის მიერ გენერირებული გარკვეული კონტენტისთვის, მათ შორის „დიფფეიკებისთვის“. აქტის 50-ე მუხლის მე-4 ნაწილის მიხედვით, „ხელოვნური ინტელექტის სისტემის განმათავსებლებმა, რომლებიც წარმოქმნიან ან მანიპულირებენ გამოსახულების, აუდიო ან ვიდეო კონტენტს, რომელიც წარმოადგენს დიფფეიკს, უნდა გაამჟღავნონ, რომ კონტენტი ხელოვნურად არის გენერირებული ან მანიპულირებული“.

აქტის მნიშვნელობა ხაზგასასმელია რამდენიმე მიმართულებით. ის წარმოშობს ვალდებულებას განხორციელდეს „დიფფეიკების“ მარკირება, ერთგვარი მითითება, რომ კონტენტი გენერირებულია ხელოვნური ინტელექტის გამოყენებით. პრობლემას წარმოადგენს „დიფფეიკის“ არა უბრალოდ არსებობა, არამედ მისი ეფექტი, რომელშიც მოიაზრება მისი უნარი გავლენა მოახდინოს ადამიანებზე, მათ აღქმებზე, შეცვალოს მათი შეხედულებები. თუ ადამიანი იჯერებს იმას, რასაც ხედავს და რაც ესმის, ყალბ, ხელოვნური ინტელექტით გენერირებულ კონტენტზე შესაბამისი მარკირება ერთ-ერთი მნიშვნელოვანი ვალდებულებაა, რომლის შესრულებას შეუძლია აღმოფხვრას ან მინიმუმამდე დაიყვანოს „დიფფეიკის“ ნეგატიური ეფექტი (Finck, 2026). ხელოვნური ინტელექტის აქტში გათვალისწინებული გამჭვირვალობის ვალდებულება მნიშვნელოვანია ე.წ. „მატყუარას დივიდენდის“ თვალსაზრისითაც. აღნიშნული გულისხმობს იმ უპირატესობას, რომელსაც „დიფფეიკ“ ტექნოლოგია აძლევს იმ პირებს, რომელთაც სურთ უარყონ მათი განცხადებების ან ქმედებების დამადასტურებელი, ავთენტური მტკიცებულებები (Chesney & Citron, 2019). როდესაც საზოგადოებაში იზრდება ცნობიერება „დიფფეიკების“ შესახებ, ადამიანები გაცილებით სკეპტიკურად უყურებენ აუდიო და ვიდეო მტკიცებულებებს (Chesney & Citron, 2019). შესაბამისად, იმ პირებს, რომლებიც დანაშაულებრივ ქმედებებში არიან ბრალდებულნი, შეუძლიათ განაცხადონ, რომ ნამდვილი ჩანაწერები წარმოადგენს „დიფფეიკს“, რაც თავის მხრივ ეჭვქვეშ აყენებს ავთენტურ მტკიცებულებებს (Chesney & Citron, 2019). გამჭვირვალობის ვალდებულება, თავისი არსით ემსახურება არა მხოლოდ კონკრეტულ ინდივიდთა ინტერესს, არამედ აღნიშნულის მიზანია საზოგადოებრივი, საჯარო ინტერესის დაცვა, როგორც ეს ასახულია ევროკავშირის „ხელოვნური ინტელექტის აქტის“ 134-ე პარაგრაფში. აქტის ოფიციალური განმარტებითა და მკვლევართა მოსაზრებით, საჯარო ინტერესი აღქმული უნდა იქნეს ფართო გაგებით, რომელშიც მოიაზრება მათ შორის საჯარო ადმინისტრაცია და სერვისები, ფუნდამენტური უფლებები (მათ შორის მართლმსაჯულების ადმინისტრირება და სამართალდამცავი ორგანოები), საზოგადოებრივი ჯანმრთელობა, გარემოს დაცვა, მომხმარებელთა უსაფრთხოება და ნებისმიერი ეკონომიკური, პოლიტიკური, სამეცნიერო ან კულტურული საკითხები (European Commission, 2026). შესაბამისად, აქტის მიზანია დაცული იქნეს საზოგადოებრივი ინტერესი ჯანმრთელობის, უსაფრთხოების და ფუნდამენტური უფლებების, მათ შორის დემოკრატიის, კანონის უზენაესობის და გარემოს დაცვის კუთხით (Gils, 2024). ევროკავშირის „ხელოვნური ინტელექტის აქტი“ ითვალისწინებს ადრეულ ეტაპზევე სამართლებრივ რეგულირებას და შესაბამისი ვალდებულებების დაკისრებას ხელოვნური

ინტელექტის სისტემების შემქმნელებისთვის, მიმწოდებლებისთვის და განმათავსებლებისთვის (Finck, 2026).

მიუხედავად რიგი ქვეყნების მიერ ეროვნულ დონეზე „დიფფეიკების“ რეგულირების სამართლებრივი მექანიზმების შემუშავების, მათ შორის 2024 წლის ევროკავშირის „ხელოვნური ინტელექტის აქტის“ არსებობისა, მზარდი ტექნოლოგიური პროცესების გათვალისწინებით, გლობალურ დონეზე კვლავაც სირთულეს წარმოადგენს „დიფფეიკების“ სამართლებრივი რეგულირება, მათი კლასიფიკაცია და საერთაშორისო სამართლის საფუძველზე მათი შეზღუდვა ან შესაბამისი რეგულირება (Santos, 2025). სამართლებრივ გამოწვევას ასევე წარმოადგენს „დიფფეიკის“ ერთიანი, საყოველთაოდ აღიარებული განმარტების არარსებობა (Santos, 2025). მკვლევართა მოსაზრებით, განმარტების ფორმულირებისას აქცენტი უნდა გაკეთდეს არა უბრალოდ „დიფფეიკის“ ტექნიკურ აღწერილობებზე, არამედ ფუნქციურ კლასიფიკაციაზე, რომელიც დაფუძნებული იქნება ზიანზე, განზრახვასა და შედეგზე, რაც შესაძლებლობას მისცემს სახელმწიფოებს ჩამოაყალიბონ რეგულირების შესაბამისი სამართლებრივი სტანდარტები (Santos, 2025). აღნიშნული გარემოების გათვალისწინებით, შეიძლება ითქვას, რომ „დიფფეიკების“ სამართლებრივი რეგულირების საშუალებები გარკვეულწილად ფრაგმენტირებულია. სახეზეა ერთიანი, უნიფიცირებული მექანიზმების ნაკლებობა.

4.4. არასამართლებრივი საშუალებები და მათი როლი „დიფფეიკების“ რეგულირებაში

როდესაც სამართლებრივი თუ არასამართლებრივი საშუალებებით ხელოვნური ინტელექტის და შესაბამისი სისტემების რეგულირებაზე ვმსჯელობთ, ბუნებრივად იბადება კითხვა იმის შესახებ, თუ რა დროითი რესურსი არსებობს ადეკვატური და ჯეროვანი რეგულირებისთვის? რამდენად პროგნოზირებადია ის, თუ რა დონეზე მოხდება ხელოვნური ინტელექტის განვითარება დროის კონკრეტულ მონაკვეთში? და, რაც მთავარია, შესაძლებელია თუ არა დროის გარკვეულ მონაკვეთში ხელოვნური ინტელექტის განვითარებისა და პროგრესის საპირწონედ, უზრუნველყოფილი იქნეს მისი ჯეროვანი რეგულირება, მათ შორის, ხელოვნური ინტელექტისა და მისი სისტემების განვითარების შედეგად წარმოშობილ გამოწვევებზე შესაბამისი რეაგირება? გაცილებით მარტივად რომ მოვახდინოთ შეკითხვის ფორმულირება - არის თუ არა ხელოვნური ინტელექტის რეგულირება იმ დონის, რაც შეესაბამება და სჭირდება მისი განვითარების დონეს?
ალბათ ვთანხმდებით, რომ ცალსახა პასუხების გაცემა აღნიშნულ შეკითხვებზე საკმაოდ რთულია. ინტერნეტის გაჩენამ დააჩქარა ტექნოლოგიური პროცესების განვითარება. ხელოვნური ინტელექტი და შესაბამისი სისტემები არნახული სისწრაფით ვითარდება, გავლენას ახდენს ჩვენს ყოველდღიურ ცხოვრებაზე, სახელმწიფოებზე, მათ ურთიერთობებზე. ხელოვნური ინტელექტი პირდაპირ თუ ირიბ გავლენას ახდენს თითქმის

ყველა სფეროზე, იქნება ეს განათლება, უსაფრთხოება, თავდაცვა, ეკონომიკა, ჯანდაცვა და სხვა.

სწორედ ამიტომ, ხელოვნური ინტელექტის რეგულირებისთვის მხოლოდ სამართლებრივი მექანიზმები არ არის საკმარისი, მით უმეტეს, იმ გარემოების გათვალისწინებით, რომ არსებული სამართლებრივი მექანიზმები ჯერ კიდევ განვითარების ადრეულ სტადიაზეა, მის შემდგომ განვითარებას არაერთი გამოწვევა ახლავს, როგორც ეროვნული, ისე საერთაშორისო სამართლის დონეზე. ამავდროულად, ტექნოლოგიები სწრაფად ვითარდება, რაც, თავის მხრივ, გავლენას ახდენს სამართლებრივი მექანიზმების შემდგომ სრულყოფასა და განვითარებაზე, რიგ შემთხვევებში, ართულებს კიდეც.

არასამართლებრივი რეგულირების მექანიზმებიდან, შეიძლება რამდენიმე გამოვყოთ. ერთ-ერთი მათგანია ე.წ. „პლატფორმის თვითრეგულირება“ (Platform Self-Regulation). აღნიშნული წარმოადგენს სხვადასხვა ინტერნეტპლატფორმის „საკუთარი პოლიტიკის წესებს“, რომელიც ნებაყოფლობითია. ასეთებია „მეტას სტანდარტები - დეზინფორმაცია და მანიპულირებული მედია“ (Meta Community Standards – Misinformation and Manipulated Media) (Meta, n.d.). მეტას პოლიტიკა ითვალისწინებს სხვადასხვა მექანიზმს, მათ შორის, ხელოვნური ინტელექტით გენერირებული ან მანიპულირებული კონტენტის მონიშვნას, მარკირებას; იმ კონტენტის წაშლას, რომელიც ეწინააღმდეგება მის სტანდარტებს. ასევე აღსანიშნავია „YouTube-ის დეზინფორმაციის პოლიტიკა“ (YouTube Misinformation Policies), და „X-ის ავთენტურობის პოლიტიკა“ (X – Authenticity Policy), რომელიც კრძალავს სინთეზურ ან მანიპულირებულ მედიას, რომელიც ასახავს მცდარ რეალობას, იწვევს პირთა დაბნეულობას და საფრთხეს უქმნის ან ზიანს აყენებს საზოგადოებრივ უსაფრთხოებას (X Corp., 2025). პოლიტიკა პლატფორმას საშუალებას აძლევს, სერიოზული დარღვევების შემთხვევაში, მონიშნოს ანგარიშები, შეზღუდოს ხილვადობა, წაშალოს კონტენტი ან შეაჩეროს ისინი (X Corp., 2025). აღნიშნულ მექანიზმებს თავისი დადებითი ეფექტი გააჩნია. თუმცა, თუ გავიზიარებთ რიგი მკვლევრების მოსაზრებას, „დიფფეიკების“ ეფექტიან რეგულირებას სჭირდება გაცილებით ფართო მიდგომა, რომელიც ითვალისწინებს ტექნიკური, ორგანიზაციული და საზოგადოებრივი ღონისძიებების ერთობლიობას და ყალიბდება სინთეზური მედიის ტექნოლოგიების განვითარებასთან ერთად (Leibowicz et al., 2021). შესაბამისად, ზემოაღნიშნული წესებისა და სტანდარტების ფუნქციონირება უნდა მოხდეს არა დამოუკიდებლად, არამედ სხვა მიდგომებთან და პრაქტიკასთან ერთად; და აღქმული უნდა იყოს როგორც საკითხის ერთიანი გადაწყვეტის შემადგენელი კომპონენტი და არა განცალკევებით ფუნქციონირებადი მექანიზმი (Leibowicz et al., 2021).

მნიშვნელოვან მექანიზმს წარმოადგენს ტექნოლოგიური კომპანიების, მედიაორგანიზაციებისა და მკვლევრების თანამშრომლობა, რომელიც ითვალისწინებს

„დიფფეიკების“ გამოვლენის მიზნით სხვადასხვა მონაცემთა ურთიერთგაზიარებას, საერთო ტექნიკური სტანდარტების შემუშავებას, ურთიერთთავსებადი იდენტიფიცირების სტადარტების შექმნას და საფრთხეების შესახებ ინფორმაციის ურთიერთგაცვლას (Reporters Without Borders, 2024; Simmons & Winograd, 2024). აღნიშნულის მაგალითს წარმოადგენს „კონტენტის წარმომავლობისა და ავთენტურობის კოალიცია“ (The Coalition for Content Provenance and Authenticity (C2PA)), რომელიც 2021 წელს ჩამოყალიბდა, მედიასა და ციფრულ სივრცეში დეზინფორმაციის საწინააღმდეგოდ (Hatchimonji, 2025). იგი უზრუნველყოფს კონტენტის სერთიფიცირებას, რომელიც ითვალისწინებს შესაბამის ფაილზე მიმაგრებულ ინფორმაციას, რომელშიც დეტალურად არის აღწერილი მისი სრული ისტორია, მათ შორის, ფაილის წარმოშობის წყარო და დროთა განმავლობაში მასში შეტანილი ყველა ცვლილება (Hatchimonji, 2025). აღნიშნული წარმოადგენს არასავალდებულო, ნებაყოფლობით მექანიზმს, რომლის ფარგლებშიც, სახელმწიფო ინსტიტუტებისგან დამოუკიდებლად, ტექნოლოგიური კომპანიები, ხელოვნური ინტელექტის დეველოპერები, მწარმოებლები, საინფორმაციო საშუალებები და მკვლევრები თანამშრომლობენ, რათა ციფრული კონტენტი უფრო გამჭვირვალე და სანდო გახადონ (Coalition for Content Provenance and Authenticity, n.d.). საორგანიზაციო კომიტეტის წევრებს წარმოადგენენ ისეთი კომპანიები, როგორებიცაა: Adobe, amazon, BBC, Google, Meta, Microsoft, OpenAI, SONY და სხვა.

არასამართლებრივი მექანიზმებიდან ასევე მნიშვნელოვანი დატვირთვა გააჩნია მედიაწიგნიერებას, რომელიც საკმაოდ ეფექტიანი მექანიზმია ონლაინკონტენტის კრიტიკულად შეფასების, ინფორმაციის სანდო წყაროების მეშვეობით გადამოწმების, მანიპულირებული მედიის ნიშნების ამოცნობისა და გენერაციული ხელოვნური ინტელექტის მუშაობის გააზრებისთვის. სამართლებრივი საშუალებებით „დიფფეიკების“ შექმნის აკრძალვის ან მათი ტექნიკური კონტროლის ნაცვლად მედიაწიგნიერება უზრუნველყოფს ადამიანებში ცნობიერების ამაღლებას და ამ გზით ამცირებს მავნე „დიფფეიკების“ ეფექტიანობას. მიზანი ამ შემთხვევაში არის ადამიანებში არა ექსპერტული ცოდნის მიღება, არამედ კრიტიკული აზროვნების და შესაბამისი უნარების გამომუშავება, რაც შეზღუდავს მანიპულირებული კონტენტის ეფექტს და მის გავრცელებას (UNESCO, n.d.). მნიშვნელოვანია ვიცოდეთ, თუ რას წარმოადგენს ხელოვნური ინტელექტი, როგორ ფუნქციონირებს ის, რა შესაძლებლობები გააჩნია მას და როგორ გამოვიყენოთ ის. რიგი მკვლევრების მოსაზრებით, „მოწინავე სინთეზური მედიის სამყაროში ხელოვნური ინტელექტის ცოდნა მხოლოდ მისი ინსტრუმენტების გამოყენებას არ გულისხმობს; იგი აუცილებელია ხელოვნური ინტელექტის მიერ ფორმირებულ რეალობაში ორიენტაციისთვის, სადაც დანახული და მოსმენილი ყოველთვის სარწმუნო აღარ არის“ (Naffi, 2025).

მიუხედავად სხვადასხვა არასამართლებრივი მექანიზმების არსებობისა, მკვლევართა მოსაზრებით, „დიფფეიკების“ წინააღმდეგ ბრძოლისა და მათი უარყოფითი ეფექტის აღმოსაფხვრელად არ არის საკმარისი მხოლოდ ერთი რომელიმე მექანიზმის არსებობა (Erokhin & Komendantova, 2026). საჭიროა მრავალმხრივი, კომპლექსური მიდგომა, რომელიც ითვალისწინებს და მოიცავს ერთდროულად როგორც ტექნოლოგიურ პროცესებს, ისე მმართველობით, ინსტიტუციურ საშუალებებს და ადამიანთა ჩართულობასა და ზედამხედველობას (Erokhin & Komendantova, 2026).

5. დასკვნა

ხელოვნური ინტელექტის განვითარება, რომელიც მზარდია, გავლენას ახდენს ადამიანთა ყოველდღიურ ცხოვრებზე. მისი ეფექტი მრავალმხრივია და ხასიათდება როგორც დადებითი, ისე უარყოფითი ტენდენციებით. ხელოვნური ინტელექტის განვითარება ქმნის ახალ შესაძლებლობებს და, ამავდროულად, ბადებს გამოწვევებს. ხელოვნურმა ინტელექტმა თავისი გავლენა მოახდინა მთლიანად საზოგადოებაზე, საზოგადოებაში არსებულ ურთიერთობებზე და ამ ურთიერთობათა მონაწილე მხარეებზე. საზოგადოებრივი ურთიერთობები თავისი შინაარსით აყალიბებს მხარეთა უფლებებსა და მოვალეობებს და იქ, სადაც იკვეთება აღნიშნული უფლებები და მოვალეობები, შესაბამისად, დგება საკითხი სამართლებრივი მოწესრიგების შესახებ. ხელოვნური ინტელექტი, მისი არსიდან გამომდინარე, საზოგადოებაში სპეციფიური ურთიერთობების ჩამოყალიბებას განაპირობებს. შესაბამისად, ალბათ ვთანხმდებით, რომ, როგორც თავად ხელოვნური ინტელექტის, ისე მის მოწინავე ტექნოლოგიებზე დამყარებული „დიფფეიკების“ როგორც სამართლებრივი, ისე არასამართლებრივი გზებითა და საშუალებებით მოწესრიგება და რეგულირება საჭიროებას წარმოადგენს.

სამართლებრივი მოწესრიგების კუთხით გამოწვევაა „დიფფეიკების“ ტრანსსასაზღვრო ბუნება, რამეთუ „დიფფეიკები“ იქმნება ერთ იურისდიქციაში, ხოლო მისი გავრცელების არეალი შესაძლოა ერთდროულად რამდენიმე სახელმწიფოს მოიცავდეს. სირთულეს წარმოადგენს ასევე ის, რომ „დიფფეიკები“ არ არის კრიმინალიზებული. ცხადია, მისი დადებითი მიზნებით გამოყენებიდან გამომდინარე, ეს ვერ მოხდება. „დიფფეიკები“ აღქმულია ტექნოლოგიის ჭრილში. მათი გამოყენებით ჩადენილი დანაშაულები და სხვა ქმედებები, როგორც წესი, უფრო ფართო კუთხით განიხილება. სირთულე ასევე უკავშირდება „დიფფეიკების“ შემქმნელთა და გამავრცელებელთა იდენტიფიცირებასა და მათ დადგენას. „დიფფეიკების“ შექმნა და მათი გავრცელება ხდება შესაბამისი ტექნოლოგიების გამოყენებით, ინტერნეტსივრცეში. სწორედ ამიტომ, ტექნოლოგიური პროცესები და საშუალებები იძლევა ასეთ პირთა ანონიმურად დარჩენის შესაძლებლობას. სამართლებრივ გამოწვევას ასევე წარმოადგენს „დიფფეიკის“ ერთიანი, საყოველთაოდ აღიარებული განმარტების არარსებობა. აღნიშნული გამოწვევების დასაძლევად,

მნიშვნელოვანი დატვირთვა გააჩნია როგორც ეროვნულ კანონმდებლობას, ისე სახელმწიფოთა შორის არსებულ სხვადასხვა საერთაშორისო ინსტრუმენტებს, რომელთა მიზანსაც სახელმწიფოთა შორის კიბერდანაშაულის წინააღმდეგ თანამშრომლობა წარმოადგენს. „დიფფეიკების“ სამართლებრივ რეგულირებასთან მიმართებით, განსაკუთრებით აღსანიშნავია 2024 წლის ევროკავშირის „ხელოვნური ინტელექტის აქტი“, რომელიც აწესებს გამჭვირვალობის მოთხოვნებს ხელოვნური ინტელექტის მიერ გენერირებული გარკვეული კონტენტისთვის, მათ შორის, „დიფფეიკებისთვის“. თუმცა, უნდა აღინიშნოს, რომ აქტი მოქმედებს ევროკავშირის მასშტაბით.

სამართლებრივ მექანიზმებთან ერთად არანაკლებ მნიშვნელოვანია არასამართებრივი საშუალებებიც. შეიძლება აღინიშნოს, რომ არასამართლებრივი საშუალებები გარკვეულწილად შემავსებელ ხასიათს ატარებენ. მნიშვნელოვანი დატვირთვა გააჩნია ე.წ. „პლატფორმის თვითრეგულირებას“, სხვადასხვა ტექნოლოგიური კომპანიების ჩართულობას, მათ ურთიერთკოორდინაციას, მედიაწიგნიერებას, საზოგადოებაში ხელოვნური ინტელექტის არსის, მისი რაობის, მის ტექნოლოგიებზე დამყარებული „დიფფეიკების“ შესახებ ცნობიერების ამაღლებას. თუმცა, არასამართლებრივ მექანიზმებს სავალდებულო ძალა არ გააჩნიათ.

მთლიანობაში, შეიძლება ითქვას, რომ „დიფფეიკების“ რეგულირებისა და მოწესრიგების საშუალებები, გარკვეულწილად, ფრაგმენტირებულია. სახეზეა ერთიანი, უნიფიცირებული მექანიზმების ნაკლებობა.

განცხადება ხელოვნური ინტელექტის გამოყენების შესახებ: სტატიის მომზადების პროცესში ხელოვნური ინტელექტი არ ყოფილა გამოყენებული.
დაფინანსება: ამ კვლევას არ მიუღია არანაირი გარე ინსტიტუციური, სახელმწიფო ან კერძო დაფინანსება.
ინტერესთა კონფლიქტი: ავტორმა განაცხადა, რომ წინამდებარე კვლევასთან დაკავშირებით, რაიმე სახის ინტერესთა კონფლიქტი არ არსებობს.

References

1. Agence France-Presse. (2023). *Research: Deepfake "news anchors" in Pro-China Footage*. Voice of America. Retrieved from www.voanews.com
2. Bird & Bird. (2024). *European Union Artificial Intelligence Act: A guide*. Retrieved from www.twobirds.com
3. Bringsjord, S., & Govindarajulu, N. S. (2018). Artificial intelligence. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 ed.). Stanford University. Retrieved from plato.stanford.edu
4. Britannica Editors. (2026, November 20). *Diffusion model*. Encyclopaedia Britannica. Retrieved from www.britannica.com
5. Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107*, 1753–1819. Retrieved from scholarship.law.bu.edu
6. Coalition for Content Provenance and Authenticity. (n.d.). *Coalition for Content Provenance and Authenticity (C2PA).* Retrieved from c2pa.org.
7. Council of Europe. (2001). *Convention on Cybercrime (ETS No. 185).* Council of Europe Treaty Series. Retrieved from www.europarl.europa.eu
8. Croitoru, F.-A., Hîji, A. I., Hondru, V., Ristea, N.-C., Irofti, P., Popescu, M., Rusu, C., Ionescu, R. T., Khan, F. S., & Shah, M. (2024). *Deepfake Media Generation and Detection in the Generative AI Era: A Survey and Outlook. arXiv*. https://doi.org/10.48550/arXiv.2411.19537
9. Department for Science, Innovation and Technology. (2025). *Online Safety Act*. GOV.UK. Retrieved from www.gov.uk
10. Detect Deepfakes. (2026). *Deepfake Law in Japan*. Retrieved from www.detectdeepfakes.com
11. Detect Deepfakes. (2026). *Deepfake Law in the Netherlands*. Retrieved from www.detectdeepfakes.com
12. Detect Deepfakes. (2026). *The Zelensky "surrender" Deepfake Video (March 2022).* Retrieved from www.detectdeepfakes.com
13. Detect Deepfakes. (n.d.). *Deepfake Laws, by country.* Retrieved from www.detectdeepfakes.com
14. Diel, A., Lalgi, T., Mellis, F. S., Teufel, A., & Bäuerle, A. (2025). The Harm of Deepfakes: A Scoping Review of Deepfakes' Negative Effects on Human Mind and Behavior. *AI & Society*. https://doi.org/10.1007/s00146-025-02774-0
15. Easttom, W. (2025). *Deepfake Technology: Emerging Threats and Security Implications. Proceedings of the 20th International Conference on Cyber Warfare and Security (ICCWS 2025), 20* (1), 79-85. https://doi.org/10.34190/iccws.20.1.3283
16. Erokhin, D., & Komendantova, N. (2026). A Review of Tools and Technologies to Combat Deepfakes. *Information, 17* (4), 347. https://doi.org/10.3390/info17040347
17. European Commission. (2026). *Draft of the Guidelines on the Implementation of the Transparency Obligations for Certain AI Systems under Article 50 of the AI Act*. European Commission. Retrieved from digital-strategy.ec.europa.eu
18. European Commission. (n.d.). *AI Act.* Shaping Europe's Digital Future. Retrieved from digital-strategy.ec.europa.eu

19. European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828.* Official Journal of the European Union, L 1689. Retrieved from eur-lex.europa.eu
20. European Union & Japan. (2018). *Strategic Partnership Agreement between the European Union and its Member States, of the one part, and Japan, of the other part*. Ministry of Foreign Affairs of Japan.
21. Finck, M. (2026). Chapter IV: Transparency Obligations for Providers and Deployers of Certain AI Systems. In *The EU Artificial Intelligence Act: A Commentary* (pp. 464-481). Oxford University Press. https://doi.org/10.1093/law/9780198925705.003.0005
22. Florkin, J. (2011, November). *Alan Turingi: Ra sheudzliat istsavlon khelovnuri intelektis shemkmnelebma gamotvlebidan, kodis gatekhvisa da gansjisgan* [Alan Turing: What artificial intelligence builders can learn from computation, codebreaking, and judgment].
23. Foreign, Commonwealth & Development Office. (2024). *Agreement between the Government of the United Kingdom of Great Britain and Northern Ireland and the Government of the United States of America on access to electronic data for the purpose of countering serious crime (TS No. 33/2024).* UK Treaty Series. Retrieved from www.gov.uk
24. Gils, T. (2024). *A detailed analysis of Article 50 of the EU's Artificial Intelligence Act*. Social Science Research Network. https://doi.org/10.2139/ssrn.4865427
25. Hatchimonji, G. (2025). *What is the Coalition for Content Provenance and Authenticity (C2PA)?* TechTarget. Retrieved from www.techtarget.com
26. Haugeland, J. (1989). *Artificial Intelligence: The Very Idea.* MIT Press. https://doi.org/10.7551/mitpress/1170.001.0001
27. He, R., Cao, J., & Tan, T. (2025). *Generative Artificial Intelligence: A Historical Perspective. National Science Review, 12* (5), nwaf050. https://doi.org/10.1093/nsr/nwaf050
28. Hunton Andrews Kurth LLP. (2024). *Understanding the EU AI Act*. Retrieved from www.hunton.com
29. Irish Council for Civil Liberties. (2025). *AI System Definition Guidelines Fail to Provide Clarity*.
30. Jejelola, F. O. (2024). The Role of Artificial Intelligence in the Eradication of Transnational Crime. *International Journal of Research and Innovation in Social Science, 8* (11), 867-882. https://doi.org/10.47772/IJRISS.2024.8110069
31. KPMG. (n.d.). *Deepfake Threats to Companies: Five Practical Steps to Address a Fast-Growing Phenomenon.* KPMG. Retrieved from kpmg.com
32. LegalClarity Team. (2026). *Information Warfare Definition: Methods and Legal Frameworks.* LegalClarity. Retrieved from legalclarity.org
33. Leibowicz, C., McGregor, S., & Ovadya, A. (2021). *The Deepfake Detection Dilemma: A Multistakeholder Exploration of Adversarial Dynamics in Synthetic Media* (arXiv:2102.06109). arXiv. https://doi.org/10.48550/arXiv.2102.06109
34. Lingvisturi Teknologiebis Jgufi. (n.d.). Intelekti [Intellect]. In *Kartuli leksikoni* [Georgian dictionary].

35. Linklaters. (2025). *EU - Guidance on "AI Systems". Does an Elephantine Definition Matter?* Retrieved from www.linklaters.com
36. Marcellino, W., Beauchamp-Mustafaga, N., Kerrigan, A., Chao, L. N., & Smith, J. (2023). *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0: Next-generation Chinese astroturfing and coping with ubiquitous AI* (PE-A2679-1). RAND Corporation. Retrieved from www.rand.org
37. Markoff, J. (2011). *John McCarthy*. *Encyclopaedia Britannica.*
38. Meta. (n.d.). *Misinformation.* Meta Transparency Center. Retrieved from transparency.meta.com
39. Mitra, A., Mohanty, S. P., & Kougianos, E. (2024). *The World of Generative AI: Deepfakes and Large Language Models*. *arXiv.* https://doi.org/10.48550/arXiv.2402.04373
40. ModelOp. (n.d.). *EU AI Act.* ModelOp. Retrieved from www.modelop.com
41. Muller, C., & Teilhard De Chardin, A. (2025). *AI Act Interpretation: Definition & Prohibitions: Consultation Feedback.* ALLAI.
42. Naffi, N. (2025). *Deepfakes and the Crisis of Knowing*. UNESCO. Retrieved from www.unesco.org
43. Nagumotu, K. (2022). Deepfakes are taking over social media: Can the law keep up? *IDEA: The Journal of the Franklin Pierce Center for Intellectual Property, 62*(2), 102–146. Retrieved from law.unh.edu
44. NATO. (2025). *NATO's approach to counter information threats*. Retrieved from www.nato.int
45. Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77
46. Nessler, B., & Wendehorst, C. (2024, December 11). *Commission guidelines on the application of the definition of an AI system and the prohibited AI practices established in the AI Act: Response of the European Law Institute*. European Law Institute. Retrieved from www.europeanlawinstitute.eu
47. Norwich University. (n.d.). *Historical examples of information warfare: From World War II and the Cold War to today*. Norwich University Online. Retrieved from online.norwich.edu
48. O'Brien, S. R. (2025, January 31). *Positive use cases of "deepfakes"*. Wilson Center. Retrieved from www.wilsoncenter.org
49. Organisation for Economic Co-operation and Development. (2024). *Explanatory memorandum on the updated OECD definition of an AI system* (OECD Artificial Intelligence Papers, No. 8). OECD Publishing. https://doi.org/10.1787/623da898-en
50. Parliament of Australia. (2024). Criminal Code Amendment (Deepfake Sexual Material) Bill 2024. Retrieved from www.aph.gov.au
51. Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society, 1*, Article 19. https://doi.org/10.1007/s44206-022-00010-6
52. Payne, L. (2026, November 20). *Deepfake*. Encyclopaedia Britannica. Retrieved from www.britannica.com
53. Popa, C., Pallath, R., Cunningham, L., Tahiri, H., Kesavarajah, A., & Wu, T. (2025). *Deepfake technology unveiled: The commoditization of AI and its impact on digital trust.* arXiv. https://doi.org/10.48550/arXiv.2506.07363

54. Premakanthan, N. (2024, August). *Information warfare through the ages: Techniques and impacts*. ResearchGate. Retrieved from www.researchgate.net
55. Reporters Without Borders. (2024). *Four recommendations for combating the threat to the right to information posed by deepfakes*. Retrieved from rsf.org
56. Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach*. Pearson.
57. Santos, E. A. (2025, December 16). *Deepfakes in international law: Legal status and gaps*. Diplomacy & Law. Retrieved from www.diplomacyandlaw.com
58. Shirish, A., & Komal, S. (2024). A socio-legal inquiry on deepfakes. *California Western International Law Journal*, *54*(2), Article 6. Retrieved from scholarlycommons.law.cwsl.edu
59. Simmons, J. C., & Winograd, J. M. (2024). *Interoperable provenance authentication of broadcast media using open standards-based metadata, watermarking and cryptography*. arXiv. https://doi.org/10.48550/arXiv.2405.12336
60. Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
61. UNESCO. (n.d.). *Media and information literacy*. Retrieved from www.unesco.org
62. United Nations Educational, Scientific and Cultural Organization. (2021, November 23). *Recommendation on the ethics of artificial intelligence*. Retrieved from www.unesco.org
63. United Nations Office on Drugs and Crime. (n.d.). *United Nations Convention against Cybercrime*. Retrieved from www.unodc.org
64. United States Congress. (2023). DEEPFAKES Accountability Act, H.R. 5586, 118th Cong. (2023–2024). Congress.gov. Retrieved from www.congress.gov
65. Vellino, A. (1986). [Review of the book *Artificial intelligence: The very idea*, by J. Haugeland]. *Artificial Intelligence*, *29*(3), 349–353. https://doi.org/10.1016/0004-3702(86)90075-5
66. Vinod, A., & Subramanian, A. (2024, September 24). Hidden dangers of AI: Deepfake technology. *Medium*. Retrieved from medium.com
67. Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Williams & Wilkins.
68. Whitaker, R. (2003). Information warfare. In D. H. Johnston (Ed.), *Encyclopedia of international media and communications* (pp. 405–415). Elsevier. Retrieved from www.sciencedirect.com
69. White House Task Force to Address Online Harassment and Abuse. (2024, May). *White House Task Force to Address Online Harassment and Abuse: Final report and blueprint*. The White House. Retrieved from bidenwhitehouse.archives.gov
70. X Corp. (2025, April). *Authenticity*. X Help Center. Retrieved from help.x.com
71. Yasar, K., Barney, N., & Wigmore, I. (2025, May 22). *What is deepfake technology?* TechTarget. Retrieved from www.techtarget.com
72. Zhou, A. (2025, May 9). Passing the Turing test. *The Classic Journal*. Retrieved from theclassicjournal.org