

## AI და საინფორმაციო ომები: ხელოვნური ინტელექტი პროპაგანდისა და დეზინფორმაციის წარმოებაში

### რეზიუმე

ხელოვნური ინტელექტი მნიშვნელოვანი ინსტრუმენტი გახდა საინფორმაციო გარემოს ფორმირებაში ახალი ტენდენციების დამსახურებით, როგორცია პერსონალიზებული დეზინფორმაციის შექმნა, ემოციური მანიპულაციის პროცესები და პოლიტიკური ნარატივების მიზანმიმართული ადაპტაცია. AI-ზე დაფუძნებული ტექნოლოგიები - ავტომატიზებული ბოტები, გენერაციული მოდელები, deepfake-ები და სხვა მეთოდები ცვლის საზოგადოებრივ აქტებს, აღძრავს იმედს ან პირიქით, შიშს. სწორედ ამ მიზეზით, საკითხის შესწავლა განსაკუთრებით აქტუალურია თანამედროვე საერთაშორისო ურთიერთობების პირობებში, სადაც ინფორმაციის კონტროლი ხშირად ისევე განსაზღვრავს ძალთა ბალანსს, როგორც სამხედრო თუ ეკონომიკური რესურსები. ბოლო წლების მოვლენები, როგორცია 2016 წლის აშშ-ს არჩევნები, COVID-19-ის „ინფორმაცია“ და 2022 წლის უკრაინის ომში გავრცელებული deepfake ვიდეოები ნათლად აჩვენებს, რომ AI არა მხოლოდ აჩქარებს დეზინფორმაციის გავრცელებას, არამედ ზრდის მის დამატებითობასა და გავლენას დემოკრატიულ სტაბილურობაზე.

საინფორმაციო ომი, როგორც ჰიბრიდული ტექტიკების გამოყენების ერთ-ერთი ხერხი საერთაშორისო ურთიერთობებში, ახალი ფენომენი არ არის, თუმცა ხელოვნური ინტელექტის სწრაფმა განვითარებამ შეცვალა პროპაგანდისა და დეზინფორმაციის გავრცელების დონე. ინფორმაციის იარაღად გამოყენება სხვადასხვა სახელმწიფოსა და არასახელმწიფოებრივი აქტორის მიერ საუკუნეების განმავლობაში ხდებოდა, მაგრამ ხელოვნური ინტელექტის გაჩენამ და განვითარებამ ახალი შესაძლებლობები და საფრთხეები გამოიწვია. თუ ადრე საინფორმაციო ომები ძირითადად კლასიკურ მედიასა და პოლიტიკური პროპაგანდის ტრადიციულ არხებზე იყო დამყარებული, დღეს მასობრივი კომუნიკაციის ცენტრალური ადგილი უკავია ციფრულ პლატფორმებს, სოციალურ მედიასა და ავტომატიზებულ სისტემებს. ეს ნაშრომი ასახავს სწორედ ამ მექანიზმებს, რომლითაც AI-სისტემები ქმნიან და ავრცელებენ ყალბ ნარატივებს, ზემოქმედებენ არჩევნებზე და ცვლიან დემოკრატიული სისტემების დინამიკას. მიუხედავად იმისა, რომ ხელოვნური ინტელექტი გამოიყენება დეზინფორმაციის წინააღმდეგ საბრძოლველადაც, ამჟამად არსებული ტექნოლოგიები ჯერ კიდევ ვერ ახერხებს მისი ეფექტიანობის მაქსიმიზაციას. სწორედ ამიტომ, მნიშვნელოვანია იმ მექანიზმების და მიდგომების ანალიზი, რომლებიც ერთი მხრივ AI-ს შესაძლებლობებს საინფორმაციო ომებში იყენებენ, ხოლო მეორე მხრივ, ცდილობენ, AI-თვე აღკვეთონ პროპაგანდისა და დეზინფორმაციის გავრცელება.

**საკვანძო სიტყვები:** ხელოვნური ინტელექტი, დეზინფორმაცია, deepfake, ინფორმაციული ომი, პროპაგანდა, კიბერუსაფრთხოება, სოციალური მედია.

## შესავალი

რათა გამოვიკვლიო, როგორ გამოიყენება ხელოვნური ინტელექტი დეზინფორმაციისა და პროპაგანდის გავრცელებაში და როგორია AI-ს მიერ გენერირებული კონტენტის გამოყენების ტენდენციები, გამოყენებული იქნება მულტიდისციპლინური მიდგომა, გაერთიანდება კონტენტ და ქეისების ანალიზი. ის მოიცავს ტექსტების, ვიზუალური მასალების, ვიდეოებისა და მომხმარებლების გამოხმაურებების შეფასებას. განსაკუთრებული ყურადღება დაეთმო იმ შინაარსს, რომელიც დაკავშირებულია ავტომატიზებული ბოტებით გავრცელებულ ინფორმაციასთან, deepfake ტექნოლოგიით შექმნილ ვიდეოებთან და AI-გენერირებულ ნარატივებთან. ნაშრომში დეტალურად გაანალიზებულია რამდენიმე შემთხვევა. თითოეული ქეისის შემთხვევაში განხილულია, როგორ იქნა გამოყენებული ხელოვნური ინტელექტი დეზინფორმაციის შესაქმნელად და გავრცელებისთვის, რა ტექნოლოგიური მექანიზმები ამოქმედდა და როგორი შედეგები გამოიღო ამ პროცესმა გლობალურ უსაფრთხოებაზე.

პირველადი და მეორეული მონაცემთა ანალიზი წარმოადგენს დამატებით ბაზას, რომელიც მხარს უჭერს კონტენტ და ქეის-კვლევას. გამოყენებულია სხვადასხვა საერთაშორისო კვლევითი ორგანიზაციის, მედიისა და აკადემიური სტატიის მიერ წარმოდგენილი მონაცემები (მაგალითად: BBC Monitoring, RAND Corporation, Graphika, StanfordInternet Observatory). გარდა ამისა, გამოყენებულია ვიზუალური მასალის ანალიზი, რომელიც უკავშირდება AI-ის მიერ გენერირებული ფოტოებისა და ვიდეოების დეტალურ შესწავლას. შეფასებულია deepfake ვიდეოების ტექნიკური მახასიათებლები, გამოსახულების რეალისტურობა და საზოგადოების ემოციურ რეაქციაზე გავლენა. ვიზუალური ანალიზი მნიშვნელოვანია იმის საჩვენებლად, როგორ მოქმედებს ფიზიკური ვიზუალის მანიპულაცია ინფორმაციის სანდოობის აღქმაზე.

## AI და სოციალური მედია

ინფორმაციულ ეპოქაში ხელოვნურმა ინტელექტმა რადიკალურად შეცვალა დეზინფორმაციის გავრცელების მასშტაბი და სიზუსტე. სოციალური მედია იქცა მთავარ ასპარეზად, სადაც ბოტები და ალგორითმები ქმნიან ფართო მხარდაჭერის ილუზიას. 2016 წლის აშშ-ს საპრეზიდენტო არჩევნებზე ათასობით ავტომატიზებულმა ანგარიშმა შეძლო პოლიტიკური დისკუსიების დომინირება და აუდიტორიის მანიპულაცია (Bradshaw & Howard 2019). ენის ბუნებრივი დამუშავების (NLP) მოდელებმა კი ტექსტების გენერირებით სანდო წყაროს იმიტაცია მოახდინეს ისე, რომ მკითხველისთვის რთული იყო სიმართლისა და ტყუილის გარჩევა (Buchanan, Lohn, Musser & Sedova 2021).

ხელოვნური ინტელექტით შექმნილი ყალბი შინაარსის თვალსაჩინო მაგალითია 2025 წლის მარტში Youtube-ზე გავრცელებული ვიდეო, სადაც წარმოდგენილი იყო ამერიკელი ევანგელისტის კენეტ კოუპლენდის გარდაცვალება. 20 წუთიანი სიუჟეტი მთლიანად გენერირებული იყო - როგორც ხმა, ისე ვიზუალური მასალა და სწრაფად გავრცელდა სოციალურ ქსელებში, ვიდრე არ გაირკვა, რომ კოუპლენდი ცოცხალი იყო და მის მიმართ არცერთი ბრალდება არ იყო წარდგენილი (Bano, Baig & Abrejo 2025).

COVID-19 პანდემიისას ბოტქსელებმა სოციალური მედიის ყველა პლატფორმაზე გაავრცელეს ანტივაქსერული გზავნილები და კონსპირაციული თეორიები, თითქოს ვირუსი ბიოლოგიური იარაღი იყო და 5G ტექნოლოგია ინვევდა დაავადებას. ასეთი ბოტები ხშირად წერდნენ კომენტარებს პოპულარულ ვიდეოებზე, რათა შეექმნათ შთაბეჭდილება, თითქოს ეს მოსაზრებები ფართოდ გავრცელებული იყო. კონტენტი მიზანმიმართულად მორგებული იყო აუდიტორიაზე: ახალგაზრდებისთვის ხაზი ესმებოდა ვაქცინის გავლენას ნაყოფიერებაზე, რელიგიური ჯგუფებისთვის - რწმენის წინააღმდეგობაზე,

მშობლებისთვის - ბავშვებისთვის საფრთხეზე. ამ პროცესმა მნიშვნელოვნად შეარყია ნდობა ოფიციალური სამედიცინო ინსტიტუტების მიმართ (Bano, Baig & Abrejo 2025).

Deepfake ტექნოლოგიამ კიდევ უფრო გააძლიერა დეზინფორმაციის დამაჯერებლობა. 2022 წელს უკრაინაში ომის დაწყებისთანავე რუსეთმა ტელეარხებზე გაავრცელა ვიდეო, რომელშიც ვოლოდიმირ ზელენსკი მოსახლეობას კაპიტულაციისკენ მოუწოდებდა. მიუხედავად იმისა, რომ ვიდეო მალევე წაიშალა, მისი გავრცელების პირველივე წუთებმა საზოგადოებაში დაბნეულობა და შიში დათესა. მსგავსი ტექნოლოგიები გამოიყენეს სხვა შემთხვევებშიც. 2024 წლის ნოემბერში სოციალურ ქსელებში გავრცელდა ვიდეოები, სადაც სერ დევიდ ბოროუს ხმის ზუსტი AI კლონი პრო-რუსულ ნარატივებით საუბრობდა. ხმის სანდოობამ და ემოციურობამ ყალბი ამბავი ბევრისთვის დამაჯერებელი გახადა. 2025 წლის მარტში TikTok-ზე გავრცელდა ვიდეო, სადაც დონალდ ტრამპი კოლუმბიის ოლქის სახელს ცვლიდა. ხმა მთლიანად გენერირებული იყო და იმეორებდა ტრამპის რეალურ ტონსა და დიქციას. ამ მხრივ, TikTok განსაკუთრებულად ეფექტიან პლატფორმად იქცა. მისი რეკომენდაციული ალგორითმები ოპტიმიზირებულია აუდიტორიის ქცევისა და ინტერესების მიხედვით, რის გამოც რადიკალური და ფსევდოპატრიოტული შინაარსი სწრაფად ვრცელდებოდა (Mazur 2022).

### AI-სთან დაკავშირებული მექანიზმები და ტექნოლოგიები

ციფრულ სივრცეში ხელოვნური ინტელექტი გადაიქცა იარაღად, რომელიც ერთდროულად ქმნის შინაარსს, მართავს მის გავრცელებას და იწვევს ემოციას. გენერაციული მოდელები - ენობრივი (GPT, LLaMA, Claude) და ვიზუალური (Stable Diffusion, DALL-E, Midjourney) - აგენერირებენ ტექსტებს, ფოტოებსა და ვიდეოებს ისეთი ხარისხით, რომ ყალბი და ავთენტური მასალების გარჩევა რთულდება (Floridi & Chiriat 2020). ამ ტექნოლოგიებს ემატება ბოტქსელები - ათასობით ავტომატიზებული პროფილი ავრცელებს ერთ ნარატივს, აჩენს „საერთო აზრის“ ილუზიას და აპროვოცირებს აზრთა დაპირისპირებას. 2020 წლის აშშ-ს არჩევნების დროს ინტერნეტ-სფეროში სწორედ ასეთი შეთქმულების თეორიებით გაჯერებული კონტენტი გავრცელდა ალგორითმებით, ხოლო კომენტარებში განსხვავებული მოსაზრება და სანინააღმდეგო არგუმენტები იცხრილებოდა (Bradshaw & Howard 2019).

ანალოგიური სიტუაციაა რეკომენდაციის სისტემებშიც: TikTok, YouTube და Facebook პრიორიტეტს ანიჭებენ ემოციურად დატვირთულ, პოლარიზებულ მასალას და ქმნიან „ბუშტებს“, სადაც მომხმარებელი წინასწარ კონსტრუირებულ რწმენებს აძლიერებს. ენის მოდელებზე დანერგილი სკრიპტები ხაზს უსვამს ფიშინგისა და სოციალური ინჟინერიის ახალ ტალღას. ტექსტები ბუნებრივ ენაზეა, სწორ ტონალობაში, ადგილობრივ მახასიათებლებსა და ინსტიტუციურ დეტალებზე (Bano, Baig & Abrejo 2025). ამიტომ მკითხველს ეჩვენება, თითქოს წერილი მოდის სანდო წყაროდან. ამას ემატება კულტურული კოდები - ისტორიული ტრავმები, იდენტობა, პოლიტიკური დაპირისპირება და უკვე მიღწეულია მიზანი: ეჭვის გაღვივება, უნდობლობის დანერგვა და საზოგადოებრივი პოლარიზაციის გაღრმავება (Gallotti 2020).

AI მექანიზმების შემდეგი საფეხური „სინთეტიკური იდენტობები“ - AI Persona Farming. იქმნება ათასობით პერსონა საკუთარი „ბიოგრაფიით“, ისტორიით, ჩვევებით. ისინი არა მარტო აზიარებენ ბმულებს, არამედ რეაგირებენ, კამათობენ, „გამოხატავენ ემოციებს“ და უბიძგებენ აუდიტორიას ერთი მიმართულებისკენ. ასეთ ქსელებს შეუძლია რამდენიმე ენაზე ერთდროულად გაავრცელოს ერთი ნარატივი. ერთ ადგილას პოსტი ჩნდება, მეორეში - კომენტარი, მესამეში ვიზუალური მასალა და ყალბი ნარატივი უკვე ფესვებს იდგამს. 2023 წლიდან მრავალ ქვეყანაში დაფიქსირდა სწორედ ასეთი შემთხვევა. ასეულობით „ახალი მომხმარებელი“ ერთდროულად ერთ თემაზე საუბრობს, ადგილობრივ სლენგს იყენებს და ადგილობრივი მედიის ტონსაც ზუსტად იმეორებს (Mazur 2022).

ზემოთ აღნიშნული გავლენის მართვას ემსახურება რეკომენდაციის ალგორითმების მანიპულაცია (RAM). მისი მიზანია ალგორითმების გამოყენება ისე, რომ კონკრეტული ნარატივი ბუნებრივად გავრცელდეს. ტექნიკა ეყრდნობა ე.წ. narrative seeding-ს: ათასობით ერთნაირი ან მსგავსი თემატიკის კონტენტი იდება სხვადასხვა პროფილიდან. ეს ვიდეოები და პოსტები შეიძლება არ შეიცავდეს თვალსაჩინოდ მცდარ ფაქტებს, ისინი ხშირად ემორისტული, დაბალი ხარისხის ან ირონიული ფორმით იწერება. თუმცა ყველა ერთსა და იმავე გზავნილს ატარებს. ასე იქმნება ვირუსულობის ილუზია: მომხმარებელი ხედავს ათეულობით მსგავს ვიდეოს, რომელთა კომენტარებშიც დომინირებს ერთი და იგივე აზრი (Riccardo 2020). შედეგად ჩნდება შთაბეჭდილება, თითქოს ეს ნარატივი საზოგადოების ფართო ფენების რეალურ განწყობას ასახავს. RAM განსაკუთრებით სახიფათოა იმიტომ, რომ ის მუშაობს საინფორმაციო „ნაცრისფერი ზონით“, სადაც სიმართლე, ირონია და ეჭვი ერთმანეთში ირევა. ამიტომ მისი გამოვლენა და გაკონტროლება პლატფორმებისთვის ბევრად უფრო რთულია.

## AI და კიბერუსაფრთხოება

ციფრულ გარემოში, სადაც დიზინფორმაცია სისწრაფითა და მასშტაბით სცდება ადამიანურ რესურსებს, ხელოვნური ინტელექტი ხდება პირველადი „სენსორი“ და „ფილტრი“. ალგორითმები რეალურ დროში აღმოაჩენს მომხმარებელთა ქცევასა და შინაარსის დინამიკას, პოულობს ნიმუშებს, რომლებიც ანომალიას ჰგავს - უცბად გახშირებულ ახალ პროფილებს, ერთსა და იმავე თემაზე სინქრონულ პოსტებს, ერთგვაროვან ენობრივ შაბლონებს (Buchanan, Lohn, Musser & Sedova 2021). იდენტურად უსაფრთხოების პლატფორმებშიც: SIEM-ში ინტეგრირებული მოდელები მრავალი წყაროდან (ლოგები, ქსელი, მონაცემები) აგროვებენ სიგნალებს, ანიჭებენ თითოეულ მოვლენას საფრთხის ქულას და წყვეტენ, საჭიროა თუ არა რეაგირება. ტექსტის დონეზე NLP ამჩნევს აგრესიულ და პოლარიზებულ რიტორიკას, აერთიანებს შეტყობინებებს და აჩვენებს, საიდან იწყება გავრცელება და სად ძლიერდება (Gallotti 2020). ტექსტში ტრანსფორმირებული მოდელები სწავლობენ სემანტიკას, ემოციურ ტონსა და ფაქტობრივ თანხვედრას; ვიზუალში და აუდიოში - კომპიუტერული ხედვა და ხმოვანი კონტროლი ამოიცნობს deepfake-ის მცირე „დაცდენებს“, არაბუნებრიობას, მიკრო-ექსპრესიებს, ტემბრის შეუთავსებლობას. ავტომატური fact-checking ადარებს ახალ ამბებს სანდო ბაზებს და აფიქსირებს შეუსაბამობასა და არაავთენტურ ქცევას (Riccardo 2020).

Auto-flagging ამოიცნობს ტრიგერებსა და შაბლონებს, ანიჭებს რისკის ქულას, ზღუდავს გავრცელებას. თუ რისკის ქულა გადააჭარბებს დადგენილ ზღვარს, სისტემა იწყებს რეაგირებას: პოსტი შეიძლება გადავიდეს დამატებითი გადამოწმების ეტაპზე, მოექცეს გამოქვეყნების წინასწარ მოდერაციაში (pre-publish moderation), ალგორითმულად შეიზღუდოს მისი გავრცელება (shadow demotion), ან სრულად გაიფილტროს და წაიშალოს პლატფორმიდან. ტენდენციების ბლოკში („emerging narratives“) სისტემა წინასწარ ამჩნევს, რა გავრცელება ვირუსულად და შეიმუშავებს შემაკავებელ ნაბიჯებს, თუმცა კულტურული მრავალფეროვნება და ენობრივი ნიუანსები მაინც რჩება ამ მხრივ სუსტ წერტილად (Floridi & Chiriatti 2020).

ბოლოს იწყება საინფორმაციო ოპერაციის სტრუქტურული დაშლა. ქსელური დარღვევის ალგორითმები პოულობს „საყრდენს“, რომლის გათიშვაც მთელ დიზინფორმაციის პროცესს ანელებს. იძულებითი იზოლაცია, ჰეშთეგების დროებითი გათიშვა, სპამის არხების მოჭრა ნარატივს სიმძაფრესა და ბილვადობას აკარგვინებს (Buchanan 2021).

## დასკვნა

თანამედროვე AI სისტემები უკვე აღჭურვილნი არიან შესაძლებლობით, შექმნან არა მხოლოდ სტატიკური პოსტები, არამედ ვიდეოები, აუდიო ჩანაწერები და კომბინირებული ნარატივები, რომლებიც მიმართულია კონკრეტული სეგმენტის ან დემოგრაფიული ჯგუფის დასარწმუნებლად. პლატფორმები უფრო მეტად იყენებენ AI-ს თავდაცვითი მექანიზმებისთვის, თუმცა ეს ტექნოლოგიები ჯერ კიდევ განვითარების პროცესშია და საჭიროებს რეგულარულ დახვეწას. ხელოვნური ინტელექტი დეზინფორმაციისა და პროპაგანდის გავრცელებაში უკვე წარმოადგენს არა მხოლოდ ინსტრუმენტს, არამედ სისტემურ იარაღს, რომელიც ცვლის ინფორმაციულ სივრცეს და გავლენას ახდენს დემოკრატიულ და საზოგადოებრივ პროცესებზე. AI-ის მიერ გენერირებული კონტენტის გამოყენების ტენდენციები მიუთითებს იმაზე, რომ მომავალში დეზინფორმაციის კამპანიები გახდება უფრო დახვეწილი და ადაპტირებული, რაც, თავის მხრივ, გაზრდის საერთაშორისო და შიდა სტაბილურობისთვის გამოწვევებს. ამის გათვალისწინებით, აუცილებელია მუდმივი დაკვირვება, მკაცრი ეთიკური ჩარჩოებისა და რეგულაციების შემოღება, რათა AI-ის გამოყენება დარჩეს სამართლიანობისა და ინფორმაციული უსაფრთხოების ფარგლებში.