ხელოვნური ინტელექტის გამოყენება საქართველოს წინააღმდეგ მიმართულ საინფორმაციო ოპერაციებთან ბრძოლაში

GCSD

ხელოვნური ინტელექტის გამოყენება საქართველოს წინააღმდეგ მიმართულ საინფორმაციო ოპერაციებთან ბრძოლაში

ავტორი: ვახტანგ ჩხენკელი

2021, თბილისი

ავტორის შესახებ:

ვახტანგ ჩხენკელი - საინფორმაციო ტექნოლოგიების სამართლის სპეციალისტი. განათლება მიღებული აქვს ტარტუსა და ვროცლავის უნივერსიტეტებში. ფლობს ილიას სახელმწიფო უნივერსიტეტის მაგისრტის ხარისხს და თბილისის სახელმწიფო უნივერსიტეტის ბაკალავრის ხარისხს სამართალმცდონეობაში. არის ლეინ კირკლანდის პეოგრამის სტიპენდიანტი და მკვლევარი. მისი საკვლევი ინტერესის საგანს წარმოადგენს: ხელოვნური ინტელექტი, ტექნოლოგიების ეთიკა და პერსონალური მონაცემების დაცვა საინფორამციო საზოგადოებაში.

დაფინანსებულია საქართველოში აშშ-ის საელჩოს ალუმნი საგრანტო პროგრამის ფარგლებში. კვლევაზე პასუხისმგებელია მისი ავტორი და შინაარსი არ შეიძლება აღქმული იყოს როგორც აშშ-ის საელჩოს ან საქართველოს სტრატეგიის და განვითარების ცენტრის პოზიცია და მოსაზრება.

Funded through the Alumni Grants Program, U.S. Embassy in Georgia. The content of this document is the sole responsibility of the author and can under no circumstances be regarded as reflecting the position of the U.S. Embassy in Georgia or Georgian Center for Strategy and Development.

დეზინფორმაცია, ყალბი ახალი ამბები და სიძულვილის ენის შემცველი გზავნილები ქართულ ინტერნეტსივრცეში გადაუჭრელ პრობლემად იქცა. ადგილობრივი და უცხოელი სბიექტები ღია და ფარული საინფორმაციო ოპერაციებით ცდილობენ საზოგადოების პოლარიზაციას, დაპირისპირების გამძაფრებას და საზოგადოების აზრის ფორმირების გზით, ამომრჩევლის გადაწყვეტილებებზე ზეგავლენას. რამდენიმე ქართული ორგანიზაცია ცდილობს, ადამიანური რესურსების გამოყენებით, გადაამოწმოს პოლიტიკოსების მიერ გაკეთებული განცხადებები და საჯარო სივრცეში გავრცელებული ინფორმაცია, თუმცა საინფორამციო ტექნოლოგიების განვითარებასთან და გზავნილების ზრდასთან ერთად, მათი გადამწომება უფრო და უფრო რთული ხდება. ხელოვნური ინტელექტის დახმარებით, ფაქტების გადამამოწმებელ ორგანიზაციებს შეუძლიათ უფრო ეფექტურად ებრძოლონ დეზინფორმაციას და ყალბ ახალ ამბებს, მონიშნონ გზავნილები და მოახდინონ შესაბამისი რეაგირება. მიუხედავად პერსპექტივისა, ხელოვნური ინტელექტი არ უნდა იქნას გაგებული როგორც საკითხის გადაჭრის უნივერსალური გზა, ისევე როგორც სხვა ტექნიკები, ხელოვნური ინტელექტიც აჩენს ეთიკურ და სამართლებრივ კითხვებს. მიუხედავად ამისა, აუცილებელია საკითხის გარშემო მსჯელობის დაწყება და ფაქტების ავტომატური გადამოწმების ეფექტური სისტემების შექმნა.

ABSTRACT

Disinformation, fake news and hate speech is rampant on the Georgian Internet, social media and forums. Local and foreign actors are trying to polarize Georgian society, create confrontation and influence public opinion and election results. Number of Georgian fact-checking organizations are manually scanning s tatements by Georgian politicians and information in public domain, however, development in information technologies and increased usage of fake news makes these techniques inefficient and burdensome. Artificial intelligence systems can facilitate more efficient fight against disinformation and fake news. Despite its advantages, artificial intelligence comes with externalities, such as ethical challenges and lawfulness. Nevertheless, it is crucial to start discussion about National artificial intelligence strategy against disinformation.

დეზინფორამცია, ყალბი ახალი ამბები და სიძულვილის ენა, უკანასკნელ წლებში, საქართველოსთვის დაუძლეველ გამოწვევად იქვა. შიდა და გარე აქტორები აქტიურად ცდილობენ დასავლური ინსტიტუტების, პოლიტიკური ოპონენტებისა და დემოკრატიული ღირებულებების დაკნინებას. მტრულად განწყობილი სახელმწიფოები და არასახელმწიფო სუბიექტებისაზოგადოებაში არსებული დაპირისპირებების, აზრთა სხვადასხვაობის, ეროვნული თუ რელიგიური სენტიმენტებისა და შიშების გამოყენებას ცდილობენ, საზოგადოებრივი აზრის ფორმირებისა და ამომრჩევლის ნების მანიპულირებისთვის. 2008 წლის კიბერ-თავდასხმის შემდეგ, საქართველოში საინფორამციო და გავლენის ოპერაციების რაოდენობამ და ხარისხმა[1] იმატა. ფეისბუქის, ფორუმებისა და საინფროამციო სააგენტოთა პლატფორმები ხშირად გამოიყენება როგორც შიდა, ისე გარე აქტორების მიერ ყალბი ახალი ამბების დეზინფორამციისა და სიძულვილის ენის შემცველი გზავნილების გასავრცელებლად. არარსებული კიბერპოლიტიკის, შესაბამისი კანონმდებლობის და ეფექტური თვითრეგულირების მექანიზმის არარსებობის პირობებში, დაინტერესებული პირები ადვილად ახერხებენ საზოგადოების პოლარიზებას და პოლიტიკური და სტრატეგიული მიზნების მიღწევას. ეს გამოწვევები განსაკუთრებით შესამჩნევია საინფორამციო ტექნოლოგიების განვითარებასთან ერთად. ტროლები, ბოტები და გავლენის აგენტები ქართული ინტერნეტსივრცის განუყოფელ ნაწილად იქცნენ. კომპიუტერული ალგორითმები უფრო და უფრო აადვილებენ დეზინფროამციისა და ყალბი ახალი ამბების გავრცელებას. მანქანური სწავლებისა და ხელოვნური ინტელექტის სისტემების გამოყენებით ტექსტის გენერირება, ვიდეოების შექმნა და ნარატივების წინასწარ შერჩეულ აუდიტორიაზე მორგება, საინფორმაციო ტექნოლოგიების განვითარებასთან ერთად, ადამიანის მიერ გენერირებული გზავნილების ხარისხს მიაღწევს, რამაც შეიძლება ქვეყნისთვის გამოუსწორებელი ზიანი მოიტანოს. დამოუკიდებელი ფაქტების გადამამოწმებელი ორგანიზაციები ცდილობენ საკუთარი რესურსებით ებრძოლონ დეზინფორამციასა და ყალბ ამბებს, ადამიანური რესურსების გამოყენებით ამოწმებენ პოლიტიკოსებისა და სხვადასხვა აქტორების მიერ გავრცელებულ ინფორამციებს და საზოგადოებას არგუმენტირებულ დასკვნებს სთავაზობენ. თუმცა, კომპიუტერული შესაძლებლობებისა და გადასამოწმებელი ინფორამციის ზრდასთან ერთად, ფაქტების ტრადიციული მეთოდით გადამოწმება უფრო და უფრო ნაკლებად ეფექტური ხდება. ხელოვნურ ინტელექტს აქვს შესაძლებლობა ხელი შეუწყოს ინტერნეტსივრცეში დეზინფორამციის, ყალბი ახალი ამბების და სიძულვილის ენის შემცვლეი გზავნილების გამოვლენას, მონიშვნასა და დეპრიორიტეტიზაციას. მანქანური სწავლების სხვადასხვა მოდელის გამოყენებით, შესაძლებელია ნაკლები დანახარჯითა და ნაკლებ დროში დიდი რაოდენობის ინფორამციის შესწავლა და შესაბამისი რეაგირება. მიუხედავად უამრავი უპირატესობისა, ხელოვნური ინტელექტი არ უნდა იქნას გაგებული როგორც დეზინფორამციასა და ყალბ ახალ ამბებთან ბრძოლის უნივერსალური საშუალება. ისევე როგორც სხვა ტექნიკებს,ფაქტების ავტომატურ გადამოწმებას აქვს ხარვეზები, რომლებიც შეიძლება გამოვლინდეს სხვადასხვა გარემოებაში. არასწორად დაგეგმილი და აღსრულებული სისტემის პირობებში, შესაძლებელია საზოგადოება მოექცეს „საინფორამციო სიბრმავის“ გარემოში, როდესაც მხოლოდ გაფილტრული და გადამუშავებული ინფორმაცია არხერხებს ადრესატამდე

[1] **Russian Cyber Attacks Against Georgia, Public Attributions and Sovereignty in Cyberspace.** Przemysław Roguski. Justsecurity.org. 06.04.20

მიღწევას, შეიზღუდოს სიტყვისა და გამოხატვის თავისუფლება, მოხდეს უმცირესობების და სხვა მოწყვლადი ჯგუფების დისკრიმინაცია. დიდ გამოწვევას წარმოადგენს სინთეზირებული გამოსახულებები და ვიდეოები, რომლებიც საქართველოში ფარული აუდიო ჩანაწერების გავრცელების აქტიური პრაქტიკის გათვალისწინებით, შესაძლებელია განსაკუთრებით ზიანისმომტანი აღმოჩნდეს ქართული დემოკრატიისათვის. ეს ნაშრომი მიზნად ისახავს დეზინფორამციასთან, ყალბ ახალ ამბებთან და სიძულვილის ენასთან ბრძოლაში ხელოვნური ინტელექტის სისტემების შესაძლებლობებისა და უპირატესობების წარმოჩენას, ამასთანავე ხაზგასმულია ის ლიმიტები, რომლებიც დაკავშირებულია ამ ტექნიკებთან. ნაშრომში ყირადღება გამახვილებულია ქართულ ინტერნეტსივრცეში გამოვლენილ არააავთენტურ ქცევაზე, ორგანიზებულ ყალბ ანგარიშებზე და მათი მოქმედების თავისებურებებზე. ნაშრომი განიხილავს ხელოვნური ინტელექტის შესაძლებლობებს, სწავლების პროცესებს და მათ გამოყენებას საქართველოს ინტერესების საზიანოდ მიმართული ქმედებების საწინააღმდეგოთ, პერსონალური მონაცემების დაცვის უგულებელყოფის რისკებს და მტრულად განწყობილი აქტორების მიერ ამ დარღვევის გამოყენების შესაძლებლობებს.

საინფორმაციო და გავლენის ოპერაციები ქართულ ინტერნეტსივრცეში

კომპიუტერული შესაძლებლობების ზრდასთან ერთად, სახელმწიფოები აქტიურად ცდილობენ, საინფორმაციო და გავლენის ოპერაციების მეშვეობით, საკუთარი პოლიტიკური, სტრატეგიული და ეკონომიკური ინტერესების გატარებას ინტერნეტსივრცეში. მრავალი ქვეყანა ავითარებს საინფორმაციო ტექნოლოგიებზე დაფუძნებულ თავდასხმით და თავდაცვით შესაძლებლობებს.[2] 2008 წლის რუსეთ-საქართველოს ომის დროს, რუსეთმა ფართომასშტაბიანი სამხედრო აგრესიის პარალელურად ქართულ ინტერნეტსივრცეს შეუტია და სახელმწიფო უწყებებისა და საინფორმაციო სააგენტოების პლატფორმების პარალიზებით შეეცადა ქვეყანაში საინფორამციო ვაკუუმი გაეჩინა და პანიკა გაემძაფრებინა. კიბერშეტევების პარალელურად, რუსეთი სრულად დაეუფლა საქართველოში შემავალ და გამავალ ინტერნეტ კავშირებს და რამდენიმე დღის განმავლობაში აკონტროლებდა ინფორმაციის მიმოსვლას.[3] ინტერნეტთან წვდომის ზრდასთან ერთად, მტრულად განწყობილ სახელმწიფოებს და არასახელმწიფო აქტორებს ეზრდებათ საინფორმაციო და გავლენის ოპერაციების გატარების ფართობი და არეალი.[4]

ქართული საზოგადოების დიდი ნაწილი არ ფლობს ინფორამციას საინფორამციო უსაფრთხოებისა და „ციფრული ჰიგიენის“ შესახებ, რაც მტრულად განწყობილი აქტორებისთვის ამარტივებს დეზინფორმაციის, ყალბი ახალი ამბებისა და სიძულვილის ენის შემცველი გზავნილების გავრცელებას. დღესდღეობით, ქართული მოსახლეობა მოწყვლადია შიდა და გარე აქტრების მიერ განხორციელებული საინფორმაციო და გავლენის ოპერაციების მიმართ. ამის ნათელი მაგალითია, საქართველოში 2018 წლის საპრეზიდენტო არჩევნების პერიოდში ფეისბუქის მომხმარებლის გიორგი ასათიანის გარშემო არსებული დისკუსია. გიორგი ასათიანი, როგორც საქართველოს პრეზიდენტობის კანდიდატი ავრცელებდა სექსისტურ და დისკრიმინაციულ გზავნილებს, რასაც ქართულ ინტერნეტსივრცეში და განსაკუთრებით სოციალურ ქსელებში მოჰყვა აქტიური განხილვა, დისკუსია და დაპირისპირება. გარკვეული პერიოდის შემდეგ გაირკვა, რომ გიორგი ასათიანის ფეისბუქის ანგარიში იყო ყალბი, ხოლო გამოყენებული პროფილის სურათი ეკუთვნოდა რუმინელ სენატორს.[5]

---

[2] მაგალითისთვის, ფეისბუქმა გამოავლინა რუსეთთან დაკავშირებული 80 000 ანგარიში, რომლებიც ინფორმაციას აწვდიდნენ 126 მილიონ ფეისბუქ მომხარებელს, დახარჯეს 100 000 ამერიკულ დოლარამდე ფეისბუქ რეკლამაში რომლებიც ჯამში 10 მილიონმა მომხარებელმა ნახა. ტვიტერის განცხადებით, რუსეთთან დაკავშირებილმა ანგარიშებმა გამოაქვეყნეს 1.4 მილიონი გზავნილი, რომლებიც ჯამში 288 ინტერაქციით. გუგლმა 2019 წელს გამოავლინა 18 იუთუბ-ანგარიში, რომლებმაც ჯამში გამოაქვეყნეს 1100 ვიდეო, ჯამში 165 000 ნახვით. **Hautala, Laura Hackers, trolls and the fight over your vote in the 2018 midterm elections What's old is new again as Election Day draws near. Here's what you need to know.** ჩინეთი აქტიურად იყენებს ბოტებს პოლიტიკური, სტრაგეგიული და კულტურული გზავნილების გასავრცელებლად, 2016 წლის კვლევის თანახმად, სოციალურ მედიაში გამოვლინდა ჩინეთის პროპაგანდისტული 488 მილიონი გზავნილი. **Meet the Chinese Trolls Pumping Out 488 Million Fake Social Media Posts New research exposes a "massive secretive operation" to fill China's internet with propaganda.**

[3] Keizer, Gregg . "Cyberattacks knock out Georgia's Internet presence". Computerworld.com 11.08.08

[4] 2008 წელს საქართველოს მოსახლეობის მხოლოდ 10%-ს ჰქონდა წვდომა ინტერნეტთან, მაჩვენებელი 2018 წელს გაიზარდა 60%-მდე. წყარო: **ინტერნეტზე წვდომის მაჩვენებელი აზერბაიჯანსა, სომხეთსა და საქართველოში.** 2020 წლის მონაცემებით ინტერნეტზე წვდომა საქართველოს მოსახლეობის 82%-ს აქვს, წყარო: **Internet penetration almost doubles in six years, 82% of households use internet in Georgia, agenda.ge, 05.06.20.** საქსტატის თანახმად, ქართველი ინტერნეტმომხამრებლის 90% ყოველდღიურად იყენებს ინტერნეტს, მათგან 95 და 54 პროცენტი ინტერნეტს სოცოალური მედიისთვის და ინფორამციის მისაღებად იყენებს. **საინფორმაციო და საკომუნიკაციო ტექნოლოგიების გამოყენება შინამეურნეობებში, საქსტატი.**

[5] **მიზოგინი პრეზიდენტობის კანდიდატი გიორგი ასათიანი არ არსებობს.** Netgazeti.ge 30.08.18

მიუხედავად ინფორამციის გადამოწმების სიმარტივისა და გაჟღერებული ინფორმაციის აშკარა პროვოკაციული შინაარსისა, ყალბი ფეისბუქის ანგარიშის მფლობელმა შეძლო ფეისბუქის ქართველი მომხმარებლების ნაწილის შეცდომაში შეყვანა და მათი ერთმანეთთან დაპირისპირება.[6] გიორგი ასათიანის შემთხვევა არასდროს გამოძიებულა, ხოლო მისი ანგარიში ამჟამად გაუქმებულია, თუმცა მსგავსი ტიპის ანგარიშები ქართულ ინტერნეტსივრცეში კვლავაც აქტიურობენ და ახერხებენ დეზინფორამციის, ყალბი ახალი ამებებისა და სიძულვილის ენის შემცველი გზავნილების გავრცელებას. „ციფრული ჰიგიენის“ უგულებელყოფის თვალსაჩინო მაგალითია ე.წ #10 წლის ჩელენჯი, რისთვისაც მოქალაქეები უცნობ ორგანიზაციას, ყოველგვარი უსაფრთხოებისა და მონაცემების დაცვის გარანტიის გარეშე, აწვდიდნენ საკუთარ ფოტოსურათებს. არსებობს მოსაზრება, რომ აღნიშნული გამოწვევა სახის ამომცნობი სისტემისათვის ინფორამციის ბაზის შესაქმნელად და ალგორითმის გასაუმჯობესებლად გამოიყენებოდა.[7]

ამერიკის შეერთებული შტატების 2016 წლის საპრეზიდენტო არჩევნების შედეგებმა, კომპიუტერული ტექნოლოგიების გამოყენებით, ამომრჩეველთა ნების მანიპულაციის მიმართ განსაკუთრებული ყურადღება გააჩინა. რუსეთთან დაკავშირებული პირები სხვადასხვა სოციალური მედიის პლატფორმებზე, ყალბი ანგარიშებით, ქმნიდნენ ჯგუფებს და აწყობდნენ დემონსტრაციებს. ისინი, ეროვნული, რელიგიური და პოლიტიკური სენტიმენტების გამოყენებით, აპირისპირებდნენ მოქალაქეებს და ახდენდნენ მათი საარჩევნო ნების მანიპულირებას.[8] მსგავსი საინფორამციო და გავლენის ოპერაციების ავტორები, საზოგადოებაში არსებულ დაპირისპირებებს, უთანხმოებებს და სენტიმენტებს იყენებენ, საჯარო და კერძო ინსტიტუტების მიმართ უნდობლობის გასაჩენად, პოლიტიკური ოპონენტების დისკრეტიდაციისა და ამომრჩევლის ნების გასაკონტროლებლად.

საქართველოს წინააღმდეგ მსგავსი ტიპის ოპერაციები როგორც შიდა, ისე გარე აქტორების მიერ[9] ხორციელდება. ასეთი ქმედებები ძირითადად რუსეთის ფედერაციიდან მოდის და საზოგადოებაში ანტიდასავლური და ქსენოფობიური განწყობის შექმნას ემსახურება, რომელიც საზოგადოებას მიეწოდება როგორც ეროვნული და პროქართული იდეოლოგია. 2006 წლის 11 მაისს, რუსეთს პრეზიდენტმა ვლადიმერ პუტინმა, მიმართვისას ყურადღება გაამახვილა რუსეთის ფედერაციის საპასუხო მოქმედებების არაპროპორციულობაზე და ინტელექტუალურ უპირატესობაზე,[11] რაცსოციალური მედიის მანიპულირებით, საინფორამციო საშუალებების კონტროლითა და გავლენის ოპერაციების სხვა მეთოდებით[11] ხორციელდება. საინფორამციო ოპერაციების ობიექტად, რუსეთის ფედერაცია როგორც საინფორმაციო ინფრასტრუქტურას, ისე ადამიანთა მიერ ინფორამციის მიღებისა და გაანალიზების პროცესს მოიაზრებს.

[6] გზავნილების ავთენტურობის გადამოწმების გარეშე, გიორგი ასათიანის განცხადებებს აქტიურად ავრცელებდნენ ქართული მედია პორტალები და ხშირ შემთხვევაში მას საქაართველოს პრეზიდენტობის კანდიდატად მოიხსენიებდნენ. ბევრი საინფორამციო პორტალი მის განცხადებებს სატყუარად (Clickbait) იყენებდნენ მკითხველის მოსაზიდად.

[7] O'Neill, Kate. **Facebook's '10 Year Challenge' Is Just a Harmless Meme—Right?** Wired.com 01.15.2019

[8] Robert S. Mueller, "United States of America v Internet Research Agency & Ors. Indictment by the Grand Jury for the District Court of Columbia."16.02.18

[9] ქართულ სახელმწიფოს არ უწაროებია სიღრმისეული გამოიძიება ქვეყანაში საინფროამციო და გავლენის ოპერაციების შესახებ, არ დადგენილან აქტორები და მათი მოქმედების ტაქტიკები. მხოლოდ ქართული არა-სამთავრობო და საინფორმაციო ორგანიზაციები აქვეყნებენ პერიოდულ ანგარიშებს ყალბი ანგარიშების, ტროლებისა და ფარული პოლიტიკური აგიტაციის შესახებ.

[10] „Наши ответы должны быть основаны на интеллектуальном превосходстве, они будут асимметричными, менее затратными“. **О доверии к власти,** 11.05.2006

[11] მეტი ინფორმაციისთვის იხილეთ: Gilles, Keir. Handbook of Russian Information Warfare. NATO defense College 2016.

მათი მიზანია, ღია და ფარული ოპერაციების მეშვეობით, ადამიანთა წარმოსახვაში მათთვის სასურველი ინფორმაციის დალექვა, რომელიც მომავალში იქონიებს გავლენას გადაწყვეტილების მიღების პროცესზე.[12] საგარეო პოლიტიკური მიზნების მისაღწევად, რუსეთს ფედერაცია სოციალურ მედიას აქტიურ ყურადღებას უთმობს. ფეისბუქის, ტვიტერისა და სხვა პლატფორმების გამოყენებით, კრემლი შეფართვით და ნაკლები დანახარჯით ახერხებს პროპაგანდის, დეზინფორმაციისა და ყალბი ახალი ამბების გავრცელებას, რომლებიც, შინაარსისა და მიზანმიმართული ნარატივების წყალობით, საზოგადოების ფართო მასებში ვრცელდება.[13] რუსეთი პროპაგანდას, დეზინფორამციასა და ყალბ ახალ ამებს ავრცელებს როგორც ტექსტურ, ისე აუდიო-ვიზუალურ პლატფორმებზე ტროლების, ბოტებისა და გავლენის აგენტების მეშვეობით.

რუსული სტილის საინფორამციო ოპერაციები მრავლად გვხვდება ქართულ ინტერნეტსივრცეში, განსაკუთრებით ფეისბუქზე. ზოგ შემთხვევაში მანიპულირება თავად სახელმწიფოს უმაღლესი თანამდებობის პირების მიერ ხდება. 2018 წლის 5 ივლისს, პრემიერ-მინისტრ გიორგი კვირიკაშვილის განცხადებას ფეისბუქზე, რომელიც ქართულ არასამთავრობო ორგანიზაციებს აკრიტიკებდა, არაბუნებრივად ბევრი მომხმარებლის გამოხმაურება მოჰყვა, რომელთა დიდ ნაწილს არაქართული ანგარიშების მფლობელები წარმოადგენდნენ.[14] პრემიერ-მინისტრის პრეს-სამსახურმა მოგვიანებით განაცხადა, რომ პრემიერ-მინისტრის ფეისბუქის გვერდზე განხორციელდა კიბერშეტევა მისი დისკრედიტაციის მიზნით, თუმცა ამის დამადასტურებელი მტკიცებულებები არ წარმოდგენილა.[15] ფეისბუქზე არააავთენტური ქცევა ასევე გამოვლინდა მამუკა ბახტაძის პრემიერობის დროსაც.[16] პროპაგანდისა და დეზინფორმაციის კამპანიაში ასევე ჩართულები არიან საქართველოს ხელისუფლებასთან დაკავშირებული პირების მიერ მართული ანგარიშები სხვადასხვა ინტერნეტ-პლატფორმებზე. 2020 წელს, ფეისბუქმა არააავთენტურ კოორდინირებულ ქცევაში ჩართული ქართული ანგარიშები გამოავლინა, რომლებიც ჩართული იყვნენ ოპოზიციისა და ხელისუფლების მიმართ კრიტიკულად განწყობილი მედიისა და არასამთავრობო ორგანიზაციების დისკრედიტაციის კამპანიაში.[17] სამართლიანი არჩევნებისა და დემოკრატიის საერთაშორისო საზოგადოების სოციალური მედიის მონიტორინგის მეორე შუალედური ანგარიში, ნათლად წარმოაჩენს ქართულ ინტერნეტსივრცეში საინფორამციო ოპერაციების მასშტაბებს და ძირითად მიმართულებებს. ანგარიშის თანახმად, ფეისბუქის გარკევული ჯგუფები, ღირებულებების თემების მანიპულირებით, საზოგადოების დამაპირი-

[12] Ajir, Media, and Bethany Vailliant. "Russian Information Warfare: Implications for Deterrence Theory." Strategic Studies Quarterly 12, no. 3 (2018): 70-89.

[13] Ibid.

[14] **"პრემიერის ბოტები" – კვირიკაშვილის ფეისბუქ მოწონებები ხელოვნურად იზრდება. ნეტგაზეთი.** 05.06.2018

[15] **პრემიერ-მინისტრის პრესსამსახური უარყოფს ე.წ. „ბოტების" გამოყენებას. რადიო თავისუფლება.** 06.06.2018

[16] **„ფეისბუქის პრემიერ"** სამართლიანი არჩევნებისა და დემოკრატიის საერთაშორისო საზოგადოება. 19.09.2019

[17] ჯამში ფეისბუქმა წაშალა 551 გვერდი, 101 ანგარიში, 122 ჯგუფი, 56 ინსტაგრამ ანგარიში. წყარო: **April 2020 Coordinated Inauthentic Behavior Report. Facebook.com 05.05.2020.** გარდა ამისა, ფეისბუქმა სხვადასხვა დროს წაშალა ნაციონალისტურ და ქსენოფობიურ ალტ-ინფოსთან დაკავშირებული 50 ანგარიში, 49 გვერდი, 4 ჯგუფი და 19 ინსტაგრამ ანგარიში, პოლიტიკურ პარტია საქართველოს პატრიოტთა ალიანსთან დაკავშირებული 54 ანგარიში, 14 გვერდი, 2 ჯგუფი და 21 ინსტაგრამ ანგარიში. ყალბი ანგარიშები განსაკუთრებით აქტიურობდნენ 2020 წლის ოქტომბრის საპარლამენტო არჩევნების წინასაარჩევნო კამპანიის პერიოდში, როდესაც მათი მეშვეობით ხდებოდა რუსეთ-საქართველოს ურთიერთობის შესახებ ინფორმაციის გავრცელება, ხელოსიფლების პოლიტიკური ოპონენტებისა და საქართველო-ნატოს პარტნიორობის დისკრედიტაცია. ყალბი ანგარიშების აქტივობები ასევე მიმართული იყო საქართველოს პატრიოტთა ალიანსისა და ქართული არჩევანის პოპულარიზაციისკენ. წყარო: **October 2020 Coordinated Inauthentic Behavior Report. Facebook 11.10.20**

სპირებელ პროპაგანდისტულ გზავნილებს ავრცელებენ, რომლებიც ქსენოფობიური, ჰომოფობიური და ანტიდასავლური გზავნილებიდან 2020 წლის საპარლამენტო არჩევნების კამპანიის პერიოდში გადაერთო პროლიტიკურ პროცესა და ამომრჩეველთა განწყობაზე ზემოქმედებაზე. ღირებულებათა საკითხებზე საზოგადოების პოლარიზების წამახალისებელი გვერდები აქტიურად იყვნენ ჩართული წინასაარჩევნო კამპანიაში, განსაკუთრებით პრორუსული აქტორების სასარგებლოდ. ანგარიშმა ასევე გამოავლინა ფეისბუქის მომხმარებლები, რომლებიც საქართველოს ხელისუფლების მადისკრედიტირებელ ინფორამციას ავრცელებენ.[18] 2019 წლიდან ფეისბუქმა დაიწყო დაფინანსებული პოლიტიკური და სოციალური გზავნილების მონიტორინგი და საზოგადოებისათვის ინფორამციის მიწოდება გზავნილების ავტორთა შესახებ.[19] მიუხედავად ამისა, მსგავსი გზავნილები, ხშირ შემთხვევაში, მრავალრიცხოვან ჯგუფებსა და გვერდებზე თავსდება, რაც მომხმარებელთა ინტერაქციის, აქტიურობისა და გაზიარების ხარჯზე ფართო მასებზე ვრცელდება, დამატებით დაფინანსებისა და გამჭვირვალობის უზრუნველყოფის გარეშე.

[18] **სოციალური მედიის მონიტორინგი - მეორე შუალედური ანგარიში.** სამართლიანი არჩევნებისა და დემოკრატიის საზოგადოება. 26.10.2020

[19] **New features will allow people to see fewer political and social issue ads on Facebook and Instagram.** Facebook.com 09.01.20

საინფორმაციო ოპერაცია და ფაქტების გადამოწმება

ინტერნეტსივრცესა და სოციალურ მედიაში გავრცელებული დეზინფორმაციისა და ყალბი ახალი ამბების გამოვლენისთვის აქტიურად მუშაობენ ფაქტების გადამამოწმებელი ორგანიზაციები. დღეისათვის საქართველოში მოქმედებს ორი ასეთი ორგანიზაცია,[20] რომლებიც ადამიანური რესურსების გამოყენებით ამოწმებენ გავრცლებულ ინფორამციას და წარადგენენ დასკვნებს მათში მოყვანილი ფაქტების შესახებ. ფაქტების გადამოწმების ეს მოდელი, დიდ ადამიანურ და ფინანსურ რესურს მოითხოვს. ამასთანავე, პრობლემას წარმოადგენს გადასამოწმებელი ინფორამციის რაოდენობა და მის მიმართ თანამშრომლის პირადი დამოკიდებულება. დეზინფორამციისა და ყალბი ახალი ამბების ზრდასთან ერთად, ინდივიდების მიერ ფაქტების გადამოწმება უფრო და უფრო ნაკლებად ეფექტურია, ხოლო, ხშირ შემთხვევაში, შეუძლებელი.

საქართველოში, ფაქტების გადამამოწმებელ ორგანიზაციებთან ერთად, გარკვეული პერიოდის განმავლობაში, მომხმარებელთა გამოხმაურებასა და შეფასებებზე დაფუძვნებული სერვისი „ვინ ვინ არის“ მოქმედებდა. ის, ინტერნეტ ბრაუზერის მეშვეობით, მომხმარებელს საშუალებას აძლევდა ფეისბუქზე აღმოჩენილი ყალბი ანგარიშების, ბოტებისა და პროსახელისუფლებო ნარატივების გამავრცელებელი ანგარიშების ილუსტრირებას. ამ ეტაპისთვის ვებ გვერდი გათიშულია, თუმცა, აპლიკაციის გადმოწერა და დაყენება კვლავაც შესაძლებელია.[21] აღნიშნული სერვისის მეშვეობით, არაერთი ყალბი ანგარიში გამოვლინდა, რომელებიც სოციალური მედიის პლატფორმებიდან მოპარულ სურათებს იყენებდნენ. მოპარული სურათების შემთხვევაში ყალბი ანგარიშის ამოცნობა და კავშირების დადგენა არ წარმოადგენს პრობლემას, თუმცა ტექნოლოგიების განვითარებასთან ერთად, კავშირების დადგენა შეუძლებელი გახდება, რამდენადაც მანქანური სწავლების ტექნიკის გამოყენებით, რეალურს მიმსგავსებული, უნიკალური გამოსახულებების გენერირებაა შესაძლებელი.

უკანასკნელ პერიოდში, ფაქტების ავტომატური გადამწომების სისტემები აქტიურად გამოიყენება, მათი ადამიანური რესურსების გამოყენებით გადამოწმების ალტერნატივად. ფაქტების ავტომატური გადამოწმება გულისხმობს ალგორითმების მეშვეობით ინფორამციის გადამოწმებას, მონიშვნას და ზოგიერთ შემთხვევაში დეპრიორიტეტიზაციას. ფეისბუქი, ტვიტერი და ციფრული სერვისების მომწოდებელი სხვა პლატფორმები, აქტიურად იყენებენ ხელოვნურ ინტელექტზე დაფუძნებული ფაქტების ავტომატურად გადამოწმების სისტემებს ტროლების, ბოტებისა და ყალბი ანგარიშების აღმოსაჩენად, შემდგომი რეაგირებისთვის. ფეისბუქის განცხადების თანახმად, ტერორიზმთან დაკავშირებული ინფორამციის 99.5%, ყალბი ანგარიშების 98.5%, სიშიშვლის შემცველი და სექსუალური ხასიათის ინფორმაციის 96% და ძალადობის შემცველი ინფორამციის 86% იშლება მონიტორინგის ავტომატური საშუალებების გამოყენებით.[22] ასევე, ფეისბუქი ხელოვნური ინტელექტის გამოყენებით სიძულვილის ენისა და ყალბი ახალი ამბების მოდერირების პროცესზე მუშაობს. ამ ეტაპისთვის, სისტემას სიძულვილის ენის შემცველი გზავნილების 38%-ის აღმოჩენა და წაშლა შეუძლია, თუმცა, ამ ეტაპზე, სისტემა ეფექტურად მხოლოდ ინგლისურ და პროტუგალიურ ენებზე ფუნქციონირებს.[23]

[20] ინფორამცია ფაქტების გადამამოწმებელი ქართული ორგანიზაციების შესახებ შეგიძლიათ იხილოთ **reporterslab.org** ვებ-გვერდზე.

[21] ვინ ვინ არის დაფუძვენულია მომხმარებლის მიერ გაზიარებულ ინფორამციაზე. საიტზე განთავსებული ინფორამციის თანამხად სერვისი რამდენიმე თვის განმავლობაში არ განახლებულა, თუმცა მუშაობის პერიოდში ის საშლებას იძლეოდა ყალბი ანგარიშების, ბოტებისა და ტროლების შესახებ ინფორამციის მიღებას.

[22] Mark Zuckerberg, “A Blueprint for Content Governance and Enforcement,” Facebook, 15.112018

[23] Koebler, J., and Cox, J. 'The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People' 23.08.2018

ფაქტების ავტომატური გადამოწმება, პერსპექტივები და პრობლემები

ხელოვნური ინტელექტი, კომპიუტერულ სისტემებში ადამიანის მსგავსი ინტელექტის სიმულირებას გულისხმობს, ალგორითმებისა და სასწავლო მონაცემების ანალიზის მეშვეობით.[24] კომპიუტერული პროგრამებისგან განსხვავებით, რომლებიც წინასწარ დადგენილ ლოგიკას და ნაბიჯებს მისდევენ, ხელოვნური ინტელექტი მანქანური სწავლებით, თავად ადგენეს მიზნის მიღწევისათვის საჭირო გზებს და ოპტიმალურ მოქმედებებს. დეზინფორამციასთან, ყალბ ახალ ამბებთან და სიძულვილის ენის შემცველ გზავნილებთან ბრძოლის კონტექსტში, ხელოვნური ინტელექტი, წინასწარ მიწოდებული სასწავლო მასალის საფუძველზე, დეზინფორამციის, მანიპულირებული ფაქტებისა და სიძულვილის ენის შემცველ გზავნილებში კავშირებისა და დამახასიათებელი ნიშან-თვისებების ანალიზს, დაკავშრებას და დადგენილი მოდელით რეაგირებას[25] გულისხმობს. შესაბამისი ალგორითმის შემუშავების შემდეგ, სისტემას მიეწოდება დიდი რაოდენობით ტექსტური ან/და აუდიო-ვიზუალური ინფორმაცია, რომელიც შედგება როგორც რეალური ფაქტების, ისე ყალბი ახალი ამბების, დეზინფორამციის, მანიპულაციებისა და სიძულვილის ენის შემცველი გზავნილებისგან. არჩეული მიდგომის შესაბამისად, სწავლების პროცესში, მეტი ან ნაკლები დოზით შეიძლება იყოს ჩართული ადამიანი.

ხელოვნური ინტელექტი, კომპიუტერულ სისტემებში ადამიანის მსგავსი ინტელექტის სიმულირებას გულისხმობს, ალგორითმებისა და სასწავლო მონაცემების ანალიზის მეშვეობით . კომპიუტერული პროგრამებისგან განსხვავებით, რომლებიც წინასწარ დადგენილ ლოგიკას და ნაბიჯებს მისდევენ, ხელოვნური ინტელექტი მანქანური სწავლებით, თავად ადგენეს მიზნის მიღწევისათვის საჭირო გზებს და ოპტიმალურ მოქმედებებს. დეზინფორამციასთან, ყალბ ახალ ამბებთან და სიძულვილის ენის შემცველ გზავნილებთან ბრძოლის კონტექსტში, ხელოვნური ინტელექტი, წინასწარ მიწოდებული სასწავლო მასალის საფუძველზე, დეზინფორამციის, მანიპულირებული ფაქტებისა და სიძულვილის ენის შემცველ გზავნილებში კავშირებისა და დამახასიათებელი ნიშან-თვისებების ანალიზს, დაკავშრებას და დადგენილი მოდელით რეაგირებას გულისხმობს. შესაბამისი ალგორითმის შემუშავების შემდეგ, სისტემას მიეწოდება დიდი რაოდენობით ტექსტური ან/და აუდიო-ვიზუალური ინფორმაცია, რომელიც შედგება როგორც რეალური ფაქტების, ისე ყალბი ახალი ამბების, დეზინფორამციის, მანიპულაციებისა და სიძულვილის ენის შემცველი გზავნილებისგან. არჩეული მიდგომის შესაბამისად, სწავლების პროცესში, მეტი ან ნაკლები დოზით შეიძლება იყოს ჩართული ადამიანი.

[24] ხელოვნური ინტელექტის ერთიანი დეფინიცია არ არსებობს. უზოგადესი განმარტებით, ხელოვნური ინტელექტი არის კომპიუტერული მეცნიერების ნაწილი, რომელიც მიწოდებულ ინფორმაციას გადაამუშავებს კომპიუტერული ალგორითმების მეშვეობით, ადგენს კავშირებს მიწდებულ ინფორმაციებს შორის, გამოკვეთს მსგავსებებს და განსხვავებებს და ახდენს რეგირებას დასახული მიზნის მისაღწევად. გამოყოფენ ხელოვნური ინტელექტის ორ სახეობას: ვიწრო (narrow AI) და ზოგადი (general or super intelligence). ვიწრო ხელოვნური ინტელექტის მაგალითია ავტომატური თარგმანი. ფაქტების ავტომატურ გადამოწმებაში სწორედ ვიწრო ხელონვური ინტელექტი გამიყენება. ზოგადი ხელოვნური ინტელექტი გულისხმობს კომპიუტერულ სისტემას, რომელსაც შეუძლია ადამიანის მსგავსად შეასრულოს მოქმედებები, აღიქვას სამყარო, გაანალიზოს ინფორმაცია და განვითარდეს საკუთარი შეხედულებს შესაბამისად. დღეისათვის ზოგადი ხელონვური ინტელექტის სისტემები არ არსებობს.

[25] ამ ნაშრომის მიზნებისთვის, ხელოვნური ინტელექტი, მანქანური სწავლება, ფაქტების ავტომატური გადამოწმება ერთი მნიშვნელობით გამოიყენება. ზოგადად, ხელოვნური ინტელექტი მოიცავს როგორც მაქნანურ სწავლებას, ისე ფაქტების ავტომატურ გადამოწმებას. ხელოვნური ინტელექტი ასევე გულისხმობს ღრმა სწავლებას, ბუნებრივი ენის დამუშავებას, მანქანურ ხედვას, ხმის ამოცნობას, რობოტიკას. დეზინფორმაციის, ყალბი ახალი ამბებისა და სიძულვილის ენის შემცველი გზავნილების მონიტორინგისთვის საჭიროა როგორც მანქანური სწავლების, ისე ღრმა სწავლებისა და ბუნებრივი ენის დამუშავების სისტემების გამოყენება. რამდენადაც დეზინფორმაცია და ყალბი ახალი ამბების გავრცელება არ შემოიფარგლება მხოლოდ ტექსტური მასალით, არამედ ხდება მანიპულირებული აუდიო-ვიზუალური გზავნილებით.

კონტროლირებადი სწავლების (supervised learning) პროცესში სისტემას მიეწოდება როგორც ყალბი ისე მანიპულირებული ინფორმაცია, რომელიც ავტორის მიერ მონიშნულია როგორც მცდარი ან ნამდვილი ინფორმაცია. სასწავლო პერიოდის გამნავლობაში, სისტემა ინფორმაციებს შორის კავშირებს აფიქსირებს, განაზოგადებს მონაცემებს და მიიღებს გადაწყვეტილებას ახალი მონაცემების ანალიზის საფუძველზე. კონტროლირებადი სწავლების (unsupervised learning) პროცესში სისტემას მიეწოდება მონაცემები წინასწარ განსაზღვრული შეფასებების გარეშე (სისტემამ არ იცის რომელი ინფორამცია არის ყალბი ან ნამდვილი). სისტემა აანალიზებს მიწოდებულ ინფორამციას, ადგენს კავშირებს და წყვეტს თუ რომელი ინფორმაცია არის ნამდვილი და რომელი ყალბი. გაძლიერებული სწავლების (reinforced learning) დროს სისტემას მიეცოდება ინფორმაცია განსაზღვრული შეფასებების გარეშე. სისტემა აანალიზებს მას, ადგენს კავშირებს და იღებს გადაწყვეტილებას ინფორამციის ნამდვილობასა და სიყალბეზე, რის შედეგადაც მასწავლებლისგან (ადამიანისგან) იღებს უკუკავშირს მიღებულ გადაწყვეტილებასთან დაკავშირებით. მანქანური სწავლების მეთოდების გამოყენებით შესაძლებელია დიდი მოცულობის ინფორამციის გადამუშავება და ანალიზი. მსგავსი ანალიზით შესაძლებელია არა მხოლოდ კავშირების დადგენა, არამედ გავრცელებული დეზინფორმაციის, ყალბი ახალი ამბებისა და სიძულვილის ენის შემცველი გზავნილების საერთო ნიშან-თვისებების აღმოჩენა. ამის შემდეგ შეიძლება ინტერნეტსივრცის ავტომატური გადამოწმება, მსგავსი ინფორამციის მონიშვნა და შემდგომი რეაგირება.[26] ეფექტური ალგორითმის შექმნასთან ერთად, აუცილებელია სისტემისთვის მაღალი ხარისხის და მრავალფეროვანი სასწავლო მასალის მიწოდება, იმისათვის რომ სისტემამ სამომავლოდ შეძლოს სხვადასხვა სახის ინფორამციის ამოცნობა და მეტი კავშირის დადგენა. სწორედ მიწოდებული ინფორმაცია წყვეტს თუ რამდენად ეფექტური და მიუკერძოებელი იქნება სისტემა. ამიტომაც მნიშვნელოვანია სასწავლო მასალის შერჩევა-მიწოდებაში მონაწილეობდეს სხვადასხვა დარგის წარმომადგენელი. წინააღმდეგ შემთხვევაში შესაძლებელია სისტემამ გაიმეოროს მასწავლებლის პირადი შეხედულებები, ცრურწმენები და ზოგ შემთხვევაში შიშები.[27]

ფაქტების ავტომატური გადამოწმების სისტემა საჭიროებს რამდენიმე საკითხის გადაწყვეტას, მათ შორისაა მისაწოდებელი ინფორამციის განმსაზღვრელი წრე და აღმოჩენილ ყალბ ინფორამციაზე რეაგირების მოდელები. იმ შემთხვევაში, თუკი სისტემის ადმინისტრატორი (მასწავლებელი) ერთპიროვნულად განსაზღვრავს მისაწოდებელი ინფორმაციის შინაარსა და მოცულობას, შესაძლებელია სისტემამ საზოგადოების ინტერესების შესაბამისად არ მიიღოს გადაწყვეტილება. ადმინისტრაციის მონოპოლიზაციის შემთხვევაში, შესაძლებელია პოლიტიკური ოპონენტების დისკრედიტაცია და მათი გზავნილების დეპრიორიტეტიზაცია. აღმოჩენილ ყალბ ინფორმაციაზე რეაგირების საუკეთესო პრაქტიკა ამჟამად არ არსებობს. დღეს მოქმედი მოდელის თანახმად, გადაწყვეტილებას გზავნილების წაშლის ან მონიშვნის შესახებ ერთპიროვნულად სერვისის მომწოდებელი იღებს. არსებობს მოსაზრებები გადაწყვეტილების აღმასრულებელი, სასამართლო ხელისუფლების ან არასამთავრობო ორგანიზაციების მონაწილეობით მიღების შესახებ.[28]

[26] მაგალითისთვის, შესაძლებელია ერთი წყაროდან მომავალი დეზინფორმაცია და ყალბი ახალი ამბები შეიცავდეს მსგავს წინადადებას, წინადადებათა წყობას, სტილისტურ და გრამატიკულ შეცდომებს და სხვა., რომლებიც აღმოჩენის შემდეგ ადვილად კონტროლირებადია შესაბამისი მეთოდების გამოყენებით.

[27] მაგალითისთვის, ისტორიულ მონაცემებზე დაფუძნებული სისტემები ხშირად ახდენენ ქალების და უმცირესობების დისკრიმინაციას.

[28] Kertysova, Katarina. Artificial Intelligence and Disinformation how Ai changes the way disinformation is produced, disseminated, and can be countered. Security and human Rights. 2018

გადაწყვეტილების მიღების მოდელთან ერთად, პრობლემას სისტემის გამჭირვალობის საკითხიც წარმოადგენს. ხელოვნური ინტელექტი განვითარების პროცესში იყენებს მანქანური სწავლების სხვადასხვა მოდელს, მათ შორის ღრმა სწავლების სისტემას (deep learning). ღრმა სწავლების სისტემა ინფორმაციას ატარებს ადამიანის ტვინის ნეირონის მსგავსი ფუნქციის მქონე ალგორითმში. ხელოვნური ნეირონული ქსელი (artificial neural network) არ იძლევა მიღებული გადაწყვეტილების ანალიზის საშუალებას, რამდენადაც გადაწყვეტილების მიღებამდე ინფორმაცია მრავალჯერ მუშავბდება მიკვლევადი მოდელის გამოყენების გარეშე. ამის გამო, შესაძლებელია მიღებული გადაწყვეტილება არ იყოს ახსნადი ან იყოს დისკრიმინაციული, ხოლო ადმინისტრატორს არ ჰქონდეს შესაძლებლობა მიაკვლიოს დისკრიმიანციის მიზეზს. სწავლების მოდელის არჩევამდე აუცილებელია ადმინისტრატორმა განიხილოს თითოეული მოდელის უპირატესობა და ნაკლი. მაგალითად, მარტივი და ახსნადი ალგორითმის შემთხვევაში, შესაძლებელია მტრულად განწყობილი აქტორების მიერ მისი მანიპულაცია ან მისი ეფექტურობის დაზიანება.[29] ფაქტების ავტომატური გადამოწმების სისტემების ერთ-ერთი უმთავრეს გამოწვევას, სიტყვისა და გამოხატვის თავისუფლებაზე მათი შესაძლო გავლენა წარმოადგენს. ადამიანისგან განსხვავებით, ხელოვნური ინტელექტის სისტემები ვერ აღიქვამენ სარკაზმს, ირონიას და სატირას, ასევე კულტურულ და სოციალურ სპეციფიკას.[30]

დეზინფორმაციისა და ყალბი ახალი ამბების ჭრილში აუცილებელია ე.წ ყალბი ვიდეოების განხილვა (Deep fakes). 2020 წლის საპარლამენტო არჩევნების დროს, ქართველმა პოლიტიკოსებმა პირველად ახსენეს მათი ყალბი ვიდეოები. ფარული აუდიოჩანაწერების გამოცდილების გათვალისწინებით, მომავალში მოსალოდნელია მათი გამოყენება პოლიტიკური ოპონენტების დისკრედიტაციისათვის. ყალბი ვიდეოები შექმნილია სპეციალური კომპიუტერული პროგრამების მეშვეობით. მაღალი ხარისხის ყალბი ვიდეოების გამოყენებით შესაძლებელია ყალბი ფაქტის რეალურად წარმოჩენა, ინსტიტუტებისა და პიროვნებების დისკრედიტაცია, ძალადობისკენ მოწოდება, საზოგადოებაში დაპირისპირების გამძაფრება და არჩევნების შედეგებზე გავლენის მოხდენა.[31] ყალბი ვიდეოების დასამზადებელი როგორც ფასიანი, ისე უფასო პროგრამები ინტერნეტშია ხელმისაწვდომი. პროგრამებით შესაძლებელია პიროვნებების ხმისა და გამოსახულების სინთეზირება და რეალურის მსგავსი მასალის შექმნა. ყალბ ვიდეოებს შეუძლიათ როგორც ყალბი ინფორამციის ნამდვილად წარმოჩენა, ისე ნამდვილი ინფორამციის წყაროს დისკრედიტაცია.[32] გამოსახულების სინთეზირება ასევე აადვილებს ინტერნეტსივრცეში არააავთენტური ანგარიშების გამოყენებას. მანქანური სწავლების მოდელით (generative adversarial netrworks) შესაძლებელია არარსებული პიროვნებების რეალისტური გამოსახულების შექმნა.[33]

[29] სისტემის დაზიანება არ არის დამოკიდებული მხოლოდ ალგორითმის გამჭირვალობის საკითხზე. მაგალითად ჩინურმა კომპანია ტენსენტმა წარმატებით შეძლო ამერიკული ავტომწარმოებლის ტესლას ავტომობილის მანქანური ხედვის სისტემაზე შეტევა, რამაც გამოიწვია ავტოპილოტის მიერ ობიექტების აღქმის აღრევა. მსგავსი შეიძლება განმეორდეს ფაქტების ავტომატური გადამოწმების პროცესშიც, როდესაც მტრულად განწყობილი სუბიექტები ახერხებენ სისტემაში შეღწევას, ალგორითმისთვის მცდარი ინფორამციის მიწოდებას, რის შედეგადაც სისტემა მოქმედებს მიღებული ინსტრუქციების თანახმად და შესაძლებელია ყალბი ინფორამცია წარმოაჩინოს ნამდვილად და პირიქით.

[30] James Vincent, "AI Won't Relieve the Misery of Facebook's Human Moderators," The Verge, 27.10 2019

[31] Robert Chesney and Danielle K. Citron, "Disinformation on Steroids: The Threat of Deep Fakes," Council on Foreign Relations (CFR), October 16, 2018, https://www.cfr.org/report/deep-fake-disinformation-steroids.

[32] Paul Chadwick, "The Liar's Dividend, and Other Challenges of Deep-Fake News," The Guardian, July 22, 2018, https://www.theguardian.com/commentisfree/2018/jul/22/deep-fake-news-donald-trump-vladimir-putin

[33] ამ მეთოდის გამოყენებისას სისტემას მიეწოდება სურათების ბაზა, სისტემა აანალიზებს სურათებს და ქმნის გამოსახულებებს, რომელთაც აქვთ სასწავლო ბაზისათვის დამახასიათებელი ნიშან-თვისებები, თუმცა არ არიან მათი იდენტური. ქართულ ინტერნეტ-სივრცეში დღეისათვის მოქმედი ყალბი ანგარიშებისათვის იყენებენ კავკასიელი მომხმარებლების ანგარიშებიდან მოპარულ სურათებს, რომელთა აღმოჩენაც შესაძლებელია ინტერნეტ ძებნის გამოყენებით. სინთეზირებული სურათების აღმოჩენა შეუძლებელია ძებნის ტრადიციული მეთოდების გამოყენებით, რამდენადაც არ აქვთ ანალოგი.

საინფორამციო ტექნოლოგიების ერაში, მონაცემები ერთ-ერთ უმნიშვნელოვანეს ღირებულებას წარმოადგენს. საინფორამციო ტექნოლოგიების დიდი ნაწილი სწორედ პერსონალური მონაცემების გამოყენებით აღწევს ადრესატებს და ნიღბავს არაავთენტურ ქცევას. პერსონალური მონაცემების დაცვა საქართველოში შესაბამისი ხარისხით არ ხორციელდება. მიუხედავად პერსონალური მონაცემების დაცვის შესახებ კანონის არსებობისა, არ არსებობს ეფექტური მექანიზმი, რომელიც მოქალაქეს საშუალებას მისცეს დაიცვას პერსონალური მონაცემები გამოყენებისა და მანიპულაციისგან. განსხვავებით ევროკავშირისგან, საქართველოში პერსონალური მონაცემების დაცვის კანონმდებლობის ეფექტური აღსრულება ვერ ხორციელდება.[34] ორგანიზაციებს შეუძლიათ თავისუფლად შეაგროვონ და გადაამუშაონ პერსონალური ინფორმაცია, მათ შორის ორგანიზაციებმა, რომლებიც შეიძლება ამ ინფორამციას არაკეთილსინდისიერი მიზნებისთვის იყენებდნენ. პერსონალური მონაცემების დამუშავება საშუალებას იძლევა შეიქმნას პიროვნების დეტალური პროფილი, რომელიც შესაძლებელია გამოყენებული იქნას არაავთენტური ქცევის დაფარვის მიზნით. დღეისათვის არაავთენტური ქცევა ადვილად ვლინდება, ყალბი ანგარიშების წერისა და კომუნიკაციის ტიპზე დაკვირვებით, თუმცა ბოტებისა და ტროლების სისტემების განვითარებასთან, პერსონალური მონაცემების შეგროვებისა და ხელოვნური ინტელექტის სისტემების გამოყენებით, შესაძლებელი იქნება ადამიანის მსგავსი ინტერაქციისა და კომუნიკაციის მოდელირება.

რუსეთიდან და ჩინეთიდან მომავალი დეზინფორამციისა და ყალბი ახალი ამბების საფრთხის გათვალისწინებით, ევროპული კომისია მუშაობს საპასუხო ზომების განხორციელების მოდელებზე.[35] ევროკავშირის ზოგიერთი წევრი ინდივიდუალურადაც აგრძელებს ბრძოლას უცხო სახელმწიფოებისგან მომდიანრე საინფორამციო საფრთხეების წინააღმდეგ. მაგალითად, 2020 წლიდან, რუსულ საინფორამციო სააგენტო „სპუტნიკს“ ესტონეთის ტერიტორიაზე საქმიანობა აეკრძალა.[36] მსგავსი აკრძალვები სხვა რუსულ პროპაგანდისტულ სააგენტოებსაც ეხება. ევროპული პარლამენტი, დეზინფროამციისა და ყალბი ახალი ამბების რეგულირების შესაძლო მოდელებზეც მუშაობს, რომლებიც მიზნად ისახავენ, ტექნოლოგიური საშუალებების გამოყენებით და სიტყვისა და გამოხატვის თავისუფლების უფლების დაცვის გათვალისწინებით, რეულირების ეფექტური მოდელის შექმნას. ინიციატივის მიხედვით, განხილულია როგორც არსებული რეგულაციების ციფრულ რეალობაზე მორგების შესაძლებლობა, ისე თანა-რეგულირება და აქტიური საკანონმდებლო ჩარევა ინტერნეტ პლატფორმების საქმიანობაში.[37] რეგულაციის პროცესთან ერთად, ინიციატივის ავტორები, ყურადღებას, მედია პლურალიზმისა და გამოხატვის თავისუფლების უზრუნველყოფის მნიშვნელობაზე ამახვილებენ. რაც, გათვალისწინებული უნდა იყოს რეგულირების შერჩეული მოდელის მიუხედავად.

[34] სახელმწიფო ინსპექტორის ვებ-გვერდზე განთავსებული გადაწყვეტილებები, როგორც წესი, ინდივიდუალურ შემთხვევებს ეხება. ვებ-გვერდზე პერსონალური მონაცემების დამუშავების სისტემური დარღვევის შესახებ გადაწყვეტილებები არ იძებნება. **ინსპექტორის გადაწყვეტილებები.** სახელმწიფო ინსპექტორის სამსახური.

[35] **Jourova: EU To Impose A 'Cost' For Spreaders Of Fake News, Such As Russia. Radio Free Europe Radio Liberty.** 03.12.20

[36] **Sputnik ends operations in Estonia.** Estonian Public Broadcasting. 01.01.20

[37] Regulating disinformation with artificial intelligence. European Parliament Research Service. 2019

დეზინფორმაცია, ყალბი ახალი ამბები და სიძულვილის ენა დემოკრადიული საზოგადოებებისათვის დიდ გამოწვევად რჩება. საინფორმაციო ტექნოლოგიების განვითარებასთან ერთად, გაიზრდება მათი რაოდენობა და ხარისხი. საქართველო აუცილებელია მოემზადოს ამ გამოწვევებისათვის და საინფორმაციო ტექნოლოგიებში ინვესტირებისა და შესაბამისი სტრატეგიის შემუშავებით, საინფორაციო თავდაცვის ეფექტური სისტემა შექმნას. ერთი რამ ცხადია, ხელოვნური ინტელექტის სისტემების განვითარების გარეშე, საქართველო ვერ გაუმკლავდება მტრულად განწყობილი აქტორების მიერ ნაწარმოებ საინფორამციო და გავლენის ოპერაციებს. საქართველოს, მეგობარი სახელმწიფოების გამოცდილების გათვალისწინებით, შეუძლია ეფექტური სტრატეგიის შექმნა და განხორციელება. ღონისძიებები შეიძლება იყოს კომპლექსური, მათ შორის მტრულად განწყობილი სახელმწიფოების მიერ მართული მედია საშუალებების კონტროლი და აკრძალვა, ეფექტური საკანონმდებლო და თვითრეგულირების მექანიზმების დანერგვა, ფაქტების ავტომატური გადამოწმების სისტემების შექმნის ხელშეწყობა და სხვა. შერჩეული მიდგომის მიუხედავად, გადაწყვეტილების მიღების პროცესში, აუცილებელია შემდეგი გარემოებების გათვალისწინება:

- დეზინფორმაცია, ყალბი ახალი ამბები და სიძულვილის ენა განსაკუთრებით ადვილად ვრცელდება მედია პლურალიზმის არარსებობის პირობებში. ამასთაანვე მნიშვნელოვანია ადამიანებს შეეძლოთ მიღებული ინფორამციის კრიტიკული ანალიზი და გადამოწმება. სხვადასხვა ქვეყნებში არსებობს დეზინფორამციისა და ყალბი ახალი ამბების გადამოწმების როგორც ეროვნული, ისე კერძო პლატფორმები. მაგალითად, ვებ გვერდი 30seconders.org ახალი ამბების გადამოწმების ეფექტურ და სწრაფ მეთოდს გვთავაზობს. პლატფორმა მკითხველს ურჩევს გაეცნოს მიღებულ ინფორამციას, შეამოწმოს ინფორამციის წყარო და მისი სანდოობა, მიუდგეს ინფორამციას კრიტიკული გონებით და გაანალიზოს გზავნილის მიზანი.[38] აუცილებელია საზოგადოებას ჰქონდეს ინფორამცია ქართულ ინტერნეტსივრცეში მოქმედი ტროლების, ბოტების, გავლენის აგენტების, მათი სტრატეგიებისა და ტაქტიკების შესახებ. საზოგადოებას უნდა მიეწოდოს ინფორმაცია, რომ სოციალურ ქსელებსა და ფორუმებე გავრცელებული გზავნილები შესაძლებელია მომდიანრეობდეს არა რეალური ადამიანების, არამედ ბოტებისა და ტროლებისგან, რომლებიც მიზანმიმართულად ცდილობენ მათ შეცდომაში შეყვანას, პოლარიზებას და დაპირისპირებას. ამ დისკუსიებში ჩართვა კი მხოლოდ გზავნილების ფართო მასებისათვის მიწვდენას ემსახურება.

- მტრულად განწყობილ სუბიექტებს, ხელოვნური ინტელექტის გამოყენებით საინფორმაციო ოპერაციების ჩასატარებლად, საქართველოს მოსახლეობის პერსონალური მონაცემები ესაჭიროებათ, ინფორმაცია მათი ქცევებისა და დამოკიდებულებების შესახებ. კიბერუსაფრთხოებისა და ციფრული ჰიგიენის უგულებელყოფის გათვალისწინებით, მტრულად განწყობილი სუბიექტები ადვილად ახერხებენ ამ ინფორამციის მოპოვებას ღია წყაროებიდან, მომხმარებლის აქტივობაზე დაკვირვებით და ანალიზით. ამას ხელს უწყობს საქართველოს კანონმდებლობის არაეფექტურობა პერსონალური მონაცემების დაცვის კუთხით. ნაკლებად სავარაუდოა, რომ საქართველომ შეძლოს პერსონალური მონაცემების

[38] ტექნიკა ყურადღებას ამახვილებს წყაროს სანდოობაზე. ხშირია შემთხვევა, როდესაც სატირული შინაარსის გზავნილები მრავალგზის გაზიარებისა და გადაწერის შემდეგ წარმოჩენილაი როგორც ფაქტი. დეზინფორამციისა და ყალბი ახალი ამბების გამავრცელებელი პირები ხშირად იყენებენ დამკვიდრებულ საინფორამციო სწყაროების ვებ-საიტებთან მიახლოებულ ინტერნეტ მისამართებს.

- დაცვის ევროპული სტანდარტის მსგავსი მიდგომების დანერგვა და აღსრულება;[39] თუმცა აუცილებელია დაიწყოს მსჯელობა თუ როგორ შეიძლება გაუმჯობესდეს არსებული ვითარება. საკანონმდებლო ცვლილებებისა და აღსრულების ეფექტური მექანიზმის შექმნამდე, შესაძლებელია პერსონალური მონაცემების დაცვა ციფრული ჰიგიენის აქტიური პოპულარიზაციით მოხდეს, როდესაც სხვადასხვა საგანმანათლებლო საფეხურზე როგორც მოსწავლეებს, ისე მასწავლებლებს მიეწოდებათ ინფორმაცია ინტერნეტში ქცევის, ინფორმაციის მიღების, გადამოწმებისა და რისკების შესახებ. ციფრული ჰიგიენისა და კიბერუსაფრთხოების კამპანიები აუცილებლად იყოს მიმართული საქართველოს ასაკოვანი მოსახლეობის მიმართ, რამდენადაც ისინი ყველაზე ნაკლებად ფლობენ კომპიუტერულ უნარებს.

- სინთეზირებული გამოსახულებისა და აუდიო ჩანაწერების წარმოების შესაძლებლობების ზრდასთან ერთად, აუცილებლად მოხდება მათი გამოყენება პოლიტიკური ოპენენტების წინააღმდეგ. შესაბამისი ტექნიკური შესაძლებლობების შექმნამდე, აუცილებელია მექანიზმის შემუშავება, რომელიც ყალბ მასალაზე დაფუძნებული კამპანიის წარმოების შესაძლებლობას შეზღუდავს. ფარული აუდიოჩანაწერების გავრცელების უსიამოვნო გამოცდილების გათვალისწინებით, შესაძლებელია საჭირო გახდეს შესაბამისი სანქციების დაწესებაზე საუბრის დაწყება, საზოგადოების ინტერესების დაცვის მიზნით.

- ტექსტური და გამოსახულებითი დეზინფორამციის, ყალბი ახალი ამბების, სიძულვილის ენის შემცველი გზავნილების გენერირების ავტომატიზაციისა და რაოდენობის ზრდა, ფაქტების ავტომატურ გადამოწმებას აუცილებელ და შეუცვლელ სერვისად აქცევს. მართალია, უახლოესი რამდენიმე წლის განმავლობაში ეს სისტემები სრულად დამოუკიდებლად ვერ შეძლებენ ფუნქციების შესრულებას, თუმცა შეუმსუბუქებენ შრომას გადამოწმების განმახორციელებელ ინდივიდებს და ორგანიზაციებს. განსხვავებით არსებული მიდგომისა, როდესაც ფაქტების კერძო და საჯარო გადამწომების სერვისი სხვადასხვა მიზანს ემსახურება, აუცილებელია ჩამოყალიბდეს ერთიანი სამოქმედო გეგმა, რომელიც შეძლებს როგორც საზოგადოების ინფორმირებას, ისე გენერირებული ინფორამციის საფუძველზე სისტემის განვითარებას და გაუმჯობესებას.

[39] მრავალი ავტორი აღნიშნავს რომ მონაცემთა დაცვის რეგულაციები ევროპაში ხელს უშლიან საინფორამციო ტექნოლოგიების შემდგომ განვითარებას, რამდენადაც ისინი ზღუდავენ ამ ტექნოლოგიებისათვის აუცილებელი მასალის, ინფორმაციაზე წვდომას(მაგალითისთვის: Zarsky, Tal. Incompatible: The GDPR in the Age of Big Data. Seton Hall Law Review. 2017) საქართველოში მკაცრი საკანონმდებლო რეგულირების პირობებში, განვითარების საწყის ეტაპზე მყოფი ბაზარი ვერ შეძლებს მსგავსი ბარიერის დაძლევას და ვერ იქნება კონკურენტული სხვა ქვეყნის საინფორამციო სერვისებთან შედარებით.

გამოყენებული ლიტერატურა

- Ajir, Media, and Bethany Vailliant. "Russian Information Warfare: Implications for Deterrence Theory." Strategic Studies Quarterly 12, no. 3 2018
- Chadwick, Paul. The Liar's Dividend, and Other Challenges of Deep-Fake News. The Guardian, 22.07.18
- Chesney, Robert and Citron, Danielle K. Disinformation on Steroids: The Threat of Deep Fakes Council on Foreign Relations. 16.10 2018
- Gilles, Keir. Handbook of Russian Information Warfare. NATO defense College 2016.
- Hautala, Laura. Hackers, trolls, and the fight over your vote in the 2018 midterm elections. 16.10.18
- Internet penetration almost doubles in six years, 82% of households use the internet in Georgia. agenda.ge. 05.06.20
- Kertysova, Katarina. Artificial Intelligence and Disinformation how Ai changes the way disinformation is produced, disseminated, and can be countered. Security and Human Rights. 2018
- Koebler, J., and Cox, J. The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People. vice.com 23.08.2018
- Muller, Robert, "United States of America v Internet Research Agency & Ors. Indictment by the Grand Jury for the District Court of Columbia."16.02.18
- O'Neill, Kate. Facebook's '10 Year Challenge' Is Just a Harmless Meme—Right?. wired.com 01.15.29
- Regulating disinformation with artificial intelligence. European Parliament Research Service. 2019
- Roguski, Przemysław. Russian Cyber Attacks Against Georgia, Public Attributions and Sovereignty in Cyberspace. Justsecurity.org 06.04.20
- Vincent, James. AI Won't Relieve the Misery of Facebook's Human Moderators. The Verge, 27.10 2019
- Wertime, David, Meet the Chinese Trolls Pumping Out 488 Million Fake Social Media Posts. foreignpolicy.com. 19.05.16
- Zuckerberg, Mark. A Blueprint for Content Governance and Enforcement. Facebook, 15.112018
- სოციალური მედიის მონიტორინგი - 2020 წლის საპარლამენტო არჩევნები. მეორე შუალედური ანგარიში. სამართლიანი არჩევნებისა და დემოკრატიის საერთაშორისო საზოგადოება. 26.10.20

ვებ რესურსები

- www.about.facebook.com
- www.geostat.ge
- www.isfed.ge
- www.netgazeti.ge
- www.news.err.ee
- www.personaldata.ge
- www.radiotavisupleba.ge
- www.redstar.ru
- www.reporterslab.org
- www.rferl.org