ხელოვნური ინტელექტის გამოყენება რაოდენობრივ კვლევებში - შესაძლებლობები და შეზღუდვები

ავტორი: რუსუდან ლომაია[1]

შესავალი

ხელოვნური ინტელექტის (AI) სწრაფმა განვითარებამ მნიშვნელოვნად გარდაქმნა აკადემიური სფერო, რაც განსაკუთრებულად აისახა სამეცნიერო კვლევის პროცესზე. დღეს მკვლევრები, კვლევის სხვადასხვა ეტაპზე, ხელოვნურ ინტელექტს აქტიურად იყენებენ - ლიტერატურის მოძიებიდან და მონაცემთა ანალიზიდან დაწყებული, ექსპერიმენტების დაგეგმვითა და სამეცნიერო ნაშრომების წერით დასრულებული. ხელოვნური ინტელექტით მხარდაჭერილი ინსტრუმენტები, რომლებიც იყენებს რობოტიკას, მანქანურ სწავლებასა (ML) და ბუნებრივი ენის დამუშავებას (NLP), მნიშვნელოვნად ამარტივებს კვლევით პროცესს სტატიების შეჯამების, შესაბამისი პუბლიკაციების რეკომენდაციისა და მკვლევრებისათვის სწორი მიმართულების ჩვენებით[2]. იქიდან გამომდინარე, რომ ხელოვნური ინტელექტი სწრაფად ვითარდება და კვლევითი მეთოდოლოგიების სხვადასხვა ეტაპში ინტეგრირდება, დადგა ახალი ცოდნის შეძენის აუცილებლობა: კონკრეტულად როგორ გამოიყენება ის აღნიშნულ მეთოდოლოგიებში, ასევე რა სარგებელი და გამოწვევები ახლავს ამ პროცესს. წინამდებარე მიმოხილვის მიზანია, ხელოვნური ინტელექტის გამოყენების შესაძლებლობების შესწავლა კონკრეტულად რაოდენობრივი კვლევის მეთოდოლოგიაში. სტატიაში განხილულია ხელოვნური ინტელექტის გამოყენების საშუალებები და მეთოდები როგორც მონაცემთა შეგროვების პროცესში, ისე მონაცემების დამუშავებისა და ანალიზის ეტაპზე.

რაოდენობრივი კვლევის განმარტება: რაოდენობრივი კვლევა მიზნად ისახავს ობიექტური თეორიების შემოწმებას გაზომვადი ცვლადებისა და რაოდენობრივი მონაცემების მეშვეობით[3].

[1] რუსუდან ლომაია ვენაში მდებარე ცენტრალური ევროპის უნივერსიტეტის (Central European University) მონაცემთა მეცნიერების საბაკალავრო პროგრამის სტუდენტია. 2026 წლის ივნისსა და ივლისში მან კვლევით ინსტიტუტ Gnomon Wise-ში ორთვიანი პრაქტიკა გაიარა, რომლის ფარგლებშიც რაოდენობრივ კვლევაში ხელოვნური ინტელექტის ინსტრუმენტების გამოყენებაზე მუშაობდა. მიმოხილვის მომზადების ფარგლებში, რუსუდან ლომაიას მენტორი იყო Gnomon Wise-ის მკვლევარი - დავით ქუტიძე.

[2] Mitra Madanchian and Hamed Taherdoost, "The Impact of Artificial Intelligence on Research Efficiency," *Results in Engineering* 26 (June 2025): 104743, https://doi.org/10.1016/j.rineng.2025.104743.

[3] John W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3. ed., [Nachdr.] (SAGE Publ, 20), https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf/

იგი ეფუძნება მონაცემებს, რომლებიც გროვდება სტრუქტურირებული და სტანდარტიზებული ინსტრუმენტების, მაგალითად, ექსპერიმენტებისა და გამოკითხვების გამოყენებით[4]. შემდგომში ამ მონაცემების საფუძველზე მოწმდება ჰიპოთეზები, იქმნება პროგნოზული მოდელები ან განისაზღვრება მიზეზ-შედეგობრივი კავშირები. ასეთი ანალიზისთვის ფართოდ გამოიყენება სტატისტიკური პროგრამები, მათემატიკური მოდელები და გამოთვლითი ალგორითმები. ანალიზის შედეგები ან ადასტურებს, ან უარყოფს არსებულ ჰიპოთეზებს, ქმნის ემპირიულად დასაბუთებულ პროგნოზებს, ან ავლენს მიზეზ-შედეგობრივ კავშირებს სხვადასხვა ცვლადს შორის. რაოდენობრივი კვლევის მთავარი მიზანია, მინიმუმამდე დაიყვანოს მიკერძოება და მიიღოს შედეგები, რომლებიც განზოგადებადია სხვადასხვა კონტექსტში[5].

დიდი ენობრივი მოდელების გამოყენება კითხვარის შედგენაში

დიდ ენობრივ მოდელებს (LLMs - Large Language Models) შეუძლიათ, დაეხმარონ მკვლევრებს კითხვარის შექმნაში. ისინი წარმოადგენს deep learning მოდელების კატეგორიას, რომლებიც დიდი მოცულობის მონაცემებზე იწვრთნება[6]. ამის შედეგად, მოდელს შეუძლია ბუნებრივი ენის და სხვა სახის კონტენტის გაგება და გენერირება დავალებების შესასრულებლად. მოდელების მაგალითებია OpenAI-ის ChatGPT, Google-ის Gemini ან Anthropic-ის Claude.

გამოკითხვაში კითხვის ფორმულირება განსაზღვრავს შეგროვებული მონაცემების ხარისხსა და ვალიდურობას. შესაბამისად, მნიშვნელოვანია, შეკითხვა არ იყოს ბუნდოვანი, არ შეიცავდეს გაუგებარ ფრაზებს ან არ აერთიანებდეს ორ პასუხს, როცა რესპონდენტს მხოლოდ ერთი პასუხის გაცემა შეუძლია (double barreled question)[7]. უმეტესწილად, კითხვარის შესამოწმებლად ტარდება კოგნიტიური ინტერვიუები და საპილოტე კვლევები, თუმცა ეს დიდ დროსა და სხვა სახის რესურსს მოითხოვს[8]. დიდ ენობრივ მოდელებს კი შეუძლიათ კითხვარის ხარისხის შეფასება

[4] Weng Marc Lim, “What Is Quantitative Research? An Overview and Guidelines,” *Australasian Marketing Journal* 33, no. 3 (2025): 325–48, https://doi.org/10.1177/14413582241264622.
[5] Lim, “What Is Quantitative Research?”
[6] Idan A. Blank, “What Are Large Language Models Supposed to Model?,” *Trends in Cognitive Sciences* 27, no. 11 (2023): 987–89, https://doi.org/10.1016/j.tics.2023.08.006.
[7] Sebastián Valenzuela et al., “Using Large Language Models for Survey Research in Communication: Opportunities and Challenges,” Communication and Change 1, no. 1 (2025): 14, https://doi.org/10.1007/s44382-025-00014-z.
[8] Divya Mani Adhikari et al., “Exploring LLMs for Automated Generation and Adaptation of Questionnaires,” *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, July 8, 2025, 1–23, https://doi.org/10.1145/3719160.3736606.

სწრაფად და ნაკლები დანახარჯით. მაგალითად, ChatGPT 4-მა წარმატებით გამოავლინა ბუნდოვანი და ორაზროვანი კითხვები, როდესაც მას მარკეტინგსა და მომხმარებელთა კვლევებში ფართოდ გამოყენებული კითხვარის შეფასება დაევალა[9].

სხვადასხვა ენაზე ან სხვადასხვა კულტურაში ჩატარებულ კვლევებში ხშირად ერთი და იმავე კითხვის განსხვავებული ინტერპრეტაციის პრობლემა დგება[10]. LLM-ის გამოყენება ამ კონტექსტშიც შესაძლებელია. მონაწილეები უკეთესად იგებენ ენობრივი მოდელების მიერ ადაპტირებულ შეკითხვებს, რაც აუმჯობესებს სხვადასხვა კულტურაში გამოყენებული კითხვარების ხარისხს[11]. მაგალითად, მკვლევრებმა გამოიყენეს გლობალური დათბობის შესახებ აშშ-ში ჩატარებული საზოგადოებრივი აზრის გამოკითხვა. დიდი ენობრივი მოდელის გამოყენებით, კითხვარი სამხრეთ აფრიკის მოსახლეობას მოარგეს. შეცვილი კითხვარი იყო ნაკლებად მიკერძოებული და მონაწილეებისთვისთვის - უფრო გასაგები[12].

LLM-ის გამოყენება ასევე შესაძლებელია არასრული კითხვების გასაუმჯობესებლად. მაგალითად, როდესაც კითხვარში კონკრეტულ კითხვაზე პასუხის რამდენიმე ვარიანტია მოცემული, მოდელს შეუძლია მკვლევარს დამატებით შესთავაზოს ისეთი პასუხი, რომელიც თავდაპირველად არ იყო გათვალისწინებული. შედეგად, შესაძლებელია უფრო სრულყოფილი კითხვარის შემუშავება[13]. მოდელებს ასევე შეუძლიათ კითხვარები უფრო დინამიკური და ადაპტირებადი გახადონ. სტატიკური კითხვარების მთავარი უარყოფითი მხარეა, რომ მონაწილეს არ აქვს შესაძლებლობა, მოითხოვოს დაზუსტება, როცა კითხვის შინაარსი მისთვის ბუნდოვანია. ჩეტბოტები შეიძლება ინტეგრირებული იყოს კითხვარის ადმინისტრირების კომპიუტერულ პროგრამებში და, საჭიროების შემთხვევაში, რესპონდენტს კითხვა სხვა ფორმულირებით მიაწოდოს, ან განუმარტოს[14].

LLM-ებს, ექსპერიმენტული სიტუაციების წარმოსაჩენად, რეალისტური სცენარების შექმნა შეუძლიათ, მაგალითად, შერჩევის პროცესის შესასწავლად გამოგონილი ორგანიზაციებისა და

[9] Sebastián Valenzuela et al., “Using Large Language Models for Survey Research in Communication: Opportunities and Challenges,” Communication and Change 1, no. 1 (2025): 14, https://doi.org/10.1007/s44382-025-00014-z.
[10] Valenzuela et al., “Using Large Language Models for Survey Research in Communication.”
[11] Valenzuela et al., “Using Large Language Models for Survey Research in Communication.”
[12] Adhikari et al., “Exploring LLMs for Automated Generation and Adaptation of Questionnaires.”
[13] Valenzuela et al., “Using Large Language Models for Survey Research in Communication.”
[14] Tara S. Behrend and Richard N. Landers, “Participant Interactions with Artificial Intelligence: Using Large Language Models to Generate Research Materials for Surveys and Experiments,” *Journal of Business and Psychology* 40, no. 6 (2025): 1275–97, https://doi.org/10.1007/s10869-025-10035-6.

კანდიდატების გენერირება[15]. ასევე შეუძლიათ დროის დაზოგვა, თუ საჭიროა არსებული კითხვარის შეცვლა და სხვა სიტუაციაზე მორგება: მაგალითად, ერთ-ერთი კვლევის მიზანი იყო საზოგადოების კმაყოფილების შესწავლა პოლიციელთა საქმიანობის მიმართ[16]. კითხვარის ნულიდან შექმნის ნაცვლად, მკვლევრებმა გამოიყენეს სამუშაოთი კმაყოფილების კვლევის უკვე არსებული ინსტრუმენტი, რომელიც დიდმა ენობრივმა მოდელმა მოარგო კონკრეტულად პოლიციის საქმიანობის კვლევის კონტექსტს. საზომი ინტრუმენტების მოდიფიცირების შემდეგ, მკვლევრებმა შეაფასეს LLM-ის მიერ გენერირებული ვარიანტები და შეარჩიეს საპილოტე ტესტირებისთვის შესაბამისი კითხვები. შედეგად, კვლევაში გამოყენებული კითხვები გაანალიზდა ფაქტორული ანალიზის მეთოდით და შეიქმნა საბოლოო კითხვების ნაკრები.

ადაპტირებული კონტენტი: დიდი ენობრივი მოდელების გამოყენება შესაძლებელია დინამიკურად, მასალის ადაპტაციისთვის, კითხვარში მონაწილის პასუხების ან ქცევის მიხედვით[17].

- მასალის მონაწილის პასუხების მიხედვით ადაპტირება, მაგალითად, მოკლე ან ბუნდოვანი პასუხის გაცემის შემთხვევაში - დამაკონკრეტებელი კითხვის დასმა;
- ადაპტიური საზომები, როცა კითხვები იცვლება რესპონდენტის პასუხებისა და გაზომილი თვისების დონის მიხედვით.

შეზღუდვები: ენობრივი მოდელების მიერ გამომუშავაბული შეკითხვები უკვე არსებულ ლიტერატურაზეა დაფუძნებული, რაც ნიშნავს, რომ არსებობს წინა კვლევებში დაშვებული შეცდომების ან მიკერძოების განმეორების რისკი. ასევე, ვინაიდან მოდელების პასუხები აქამდე გამოცემულ კვლევებს ან მონაცემებს ეფუძნება, ისინი ყოველთვის არ ასახავს უახლეს ინფორმაციას, ან არ ითვალისწინებს ყველა კულტურულ ან დემოგრაფიულ ნიუანსს, განსაკუთრებით, როცა ეს ინფორმაცია სწრაფად იცვლება[18]. ასევე, მოთხოვნის (prompt) სხვადასხვანაირად ფორმულირების შემთხვევაში, მოდელმა შეიძლება მკვლევარს სრულიად სხვადასხვა პასუხი გასცეს. შესაბამისად, მნიშვნელოვანია, მკვლევარი კრიტიკულრად აფასებდეს მოდელის მიერ წარმოებულ შეკითხვებს/ინფორმაციას.

[15] Behrend and Landers, “Participant Interactions with Artificial Intelligence.”
[16] Behrend and Landers, “Participant Interactions with Artificial Intelligence.”
[17] Behrend and Landers, “Participant Interactions with Artificial Intelligence.”
[18] Valenzuela et al., “Using Large Language Models for Survey Research in Communication.”

ხელოვნური ინტელექტი, როგორც სატელეფონო კითხვარის ინტერვიუერი

გამოკითხვა სხვადასხვა ფორმატში ტარდება: შეიძლება ჩატარდეს ონლაინ, ფოსტით, პირისპირ ან სატელეფონო გამოკითხვის სახით. სატელოფონო გამოკითხვას თავისი დადებითი მხარეები აქვს, კერძოდ, ონლაინ ან ფოსტით ჩატარებულ კითხვართან შედარებით, სრული პასუხის მიღების უკეთესი გარანტია და ყველა ასაკობრივი ჯგუფის ჩართვა (უფროს ასაკობრივ ჯგუფებს შეიძლება გაუჭირდეთ ონლაინ კითხვარის შევსება, ტექნოლოგიის გამოყენების უნარის ნაკლებობის გამო). თუმცა სატელეფონო გამოკითხვების ჩატარებას დიდი დრო და სხვა სახის რესურსი სჭირდება, ასევე, არ ითვალისწინებს ინტერვიუერის მიკერძოების რისკს. დიდ ენობრივ მოდელებს შეუძლიათ ჩაატარონ გამოკითხვა და შეამცირონ მკვლევრების მიერ დახარჯული რესურსი[19].

ენობრივი მოდელის ინტერვიუერის როლის შესაფასებლად, ერთ-ერთ კვლევაში კითხვარი მომზადდა REDcap-ით - ვებგვერდით, რომელიც განკუთვნილია კითხვარების შექმნისა და მონაცემების შენახვისთვის. BlandAI-ის მიერ შექმნილი დიდი ენობრივი მოდელის ინტერვიუერი რეკავდა კითხვარის მონაწილეებთან და ინახავდა როგორც კითხვარის აუდიოჩანაწერს, ისე დიალოგის ტრანსკრიპტს. OpenAI-ს Chatgpt-4o (11.2024-ის ვერსია) აანალიზებდა დიალოგს და შედეგებს ტვირთავდა REDcap-ში. გამოკითხვა ინგლისურ ენაზე ჩატარდა[20].

დიდი ენობრივი მოდელის ეფექტურობა სამნაირად გაიზომა. BlandAI-ს ინტერვიუერის ტრანსკრიფციის უნარი, კითხვარის ანალიზი chatGPT 4o-ის ტრანსკრიპტების მიხედვით, გამოკითხვაში მონაწილეების გამოცდილება დიდი ენობრივი მოდელის აგენტთან.

ხელოვნური ინტელექტი უმეტესწილად აუდიოტრანსკრიფციას მაღალი სიზუსტით აკეთებს (6.4% WER - word error rate-ით, როცა 5-10% კარგი შედეგის ინდიკატორია). ChatGPT-ი შეკითხვებზე პასუხებს აანალიზებს ან სრულყოფილად სწორად, ან საშუალოდ 97%-იანი წარმატებით. კითხვარში მონაწილეების უმეტესობის შეფასებით, აგენტი ეფექტურად ხსნიდა კითხვარის მიზანს და საკმაოდ კეთილგანწყობილი იყო[21].

[19] Kurmanbek Kaiyrbekov et al., “Automated Survey Collection with LLM-Based Conversational Agents,” *JAMIA Open* 8, no. 5 (2025): ooaf103, https://doi.org/10.1093/jamiaopen/ooaf103.
[20] Kaiyrbekov et al., “Automated Survey Collection with LLM-Based Conversational Agents.”
[21] Kaiyrbekov et al., “Automated Survey Collection with LLM-Based Conversational Agents.”

სხვა კვლევაში ხელოვნურმა ინტელექტმა ჩაატარა SSRS-ის სტანდარტული საზოგადოებრივი აზრის გამოკითხვა სატელეფონო ინტერვიუს მეთოდით. მონაწილეების დაახლოებით 80%-მა აღნიშნა, რომ გამოკითხვაში მონაწილეობის გამოცდილება იყო დადებითი/ნეიტრალური, მონაწილეების 70%-ისთვის მოდელის კითხვები საკმაოდ გასაგები აღმოჩნდა, ხოლო 60%-მა ჩათვალა, რომ მოდელმა მათი პასუხები გაიგო[22].

შეზღუდვები: ChatGPT 4o-ს გარკვეული ტიპის პასუხების ანალიზი უფრო უჭირდა, მაგალითად, გამოკითხვის მონაწილის შემოსავალის შესახებ პასუხის სწორად ანოტირება, ასევე BlandAI-ს მოდელი უკეთესად აკეთებდა იმ მონაწილეების პასუხების ტრანსკრიპციას, რომელთა მშობლიურ ენაზეც ჩატარდა გამოკითხვა. ხმოვანი აგენტი სრულიად სწორად ვერ ახერხებდა ტრანსკრიპციას მეორე კვლევაშიც. მონაწილეების ნაწილმა კი აგენტი მოიხსენია, როგორც “რობოტული”.

მონაცემთა შეგროვების პროცესი TESA ხელოვნური ინტელექტის ალგორითმის საშუალებით

სამეცნიერო კვლევაში მონაცემთა შეგროვების ტიპური პროცესი გულისხმობს, რომ კვლევაში მონაწილე ჯგუფები მაქსიმალურად დაბალანსებულად იყვნენ წარმოდგენილი. მაგალითად, ექსპერიმენტის თითოეულ ეტაპზე ადამიანების თანაბარ განაწილებას. ეს მიდგომა ლოგიკურია, თუმცა ზოგ შემთხვევაში ყველა მონაცემს ერთნაირი მნიშვნელობა არ ენიჭება: კონკრეტული მონაცემი შეიძლება საინტერესო ანომალიას წარმოადგენდეს და უფრო სიღრმისეულად გამოკვლევას საჭიროებდეს. თუმცა ანომალიის ხელით გამოკვლევა დიდ დროს მოითხოვს, განსაკუთრებით, როცა მონაცემების დიდი რაოდენობა სწრაფად გროვდება.

მკვლევრებმა განავითარეს ხელოვნური ინტელექტის ალგორითმი TESA (Threshold Estimating Sampling Algorithm). მისი მიზანია მონაცემთა შეგროვების ოპტიმიზაცია, მკვლევარსა და ალგორითმს შორის უკუკავშირის შექმნით. საწყისი მონაცემების შეგროვებისა და შესწავლის შემდეგ, მკვლევარი ალგორითმს მიმართავს საინტერესო ანომალიებისკენ (რომლებსაც საკვანძო წერტილებს - key points - უწოდებენ)[23]. ალგორითმი საკვანძო წერტილებს, დამატებითი შერჩევის გზით, უფრო დეტალურად იკვლევს. სისტემას ორი მიზანი აქვს: ერთი მხრივ, ის უფრო

[22] Danny D. Leybzon et al., “AI Telephone Surveying: Automating Quantitative Data Collection with an AI Interviewer,” version 1, preprint, arXiv, 2025, https://doi.org/10.48550/ARXIV.2507.17718.
[23] Travis Mandel et al., “AI-Assisted Scientific Data Collection with Iterative Human Feedback,” *Proceedings of the AAAI Conference on Artificial Intelligence* 35, no. 7 (2021): 5957–66, https://doi.org/10.1609/aaai.v35i7.16744.

სიღრმისეულად იკვლევს თავდაპირველად განსაზღვრულ საკვანძო წერტილებს, ხოლო მეორე მხრივ, აგრძელებს სხვა კატეგორიების მონაცემთა შეგროვებას, ახალი პოტენციური საკვანძო წერტილების აღმოსაჩენად.

ხელოვნური მონაცემების გენერირება

მაღალი ხარისხის მონაცემების ბაზა კარგი კვლევის აუცილებელი კომპონენტია. მაგალითად, სამედიცინო სფეროში, დაავადების დეტექციისთვის, სპეციალური შეკვეთით დამზადებული წამლების გაუმჯობესებისა და მკურნალობის შეფასებისთვის, საჭიროა დიდი რაოდენობის პაციენტთა მონაცემები[24]. ფინანსებში ტრანზაქციების მონაცემების ანალიზი მკვლევრებს რისკის შეფასებაში, თაღლითობის დეტექციასა და ალგორითმული ვაჭრობის სტრატეგიების დახვეწაში ეხმარება[25].

მკვლევრებს ხვდებათ დაბრკოლებები მონაცემების შეგროვების დროს. ორგანიზაციებისთვის ზუსტი მონაცემების წყაროების პოვნა რთულია, ხოლო სხვისი მონაცემების მოპოვება ხშირად არაერთ ბიუროკრატიულ ბარიერს აწყდება. მონაცემთა ხარისხისა და მოპოვების სირთულის გარდა, მკვლევრებს ეთიკის საკითხიც აბრკოლებთ. მომხმარებლის მონაცემთა კონფიდენციალურობის დაცვიდან გამომდინარე, ბაზებზე წვდომა ხშირად შეუძლებელია. ამ სირთულეების დასაძლევად ერთ-ერთი ვარიანტია ხელოვნური მონაცემების გენერირება, რაც მკვლევრებს აძლევს საშუალებას, შექმნან რეალისტური, თუმცა ფიქციური მონაცემები, რომლებიც სენსიტიურ ინფორმაციას არ ამჟღავნებს.

ხელოვნური ინტელექტის საშუალებით გამომუშავებული მონაცემების ერთ-ერთი უპირატესობაა მათი გამოყენება სხვადასხვა სფეროში. სამედიცინო ან ფინანსურ სექტორში ხშირად რთულია დიდი მოცულობის ანოტირებული მონაცემების შოვნა. ხელოვნურ მონაცემებს ამ სირთულის გადაჭრა დამატებითი სასწავლო მაგალითების შექმნით შეუძლია. გარდა ამისა, ხელოვნური მონაცემები გამოიყენება მონაცემთა გაფართოების (data augmentation) პროცესში, რადგან ისინი მოდელებს უფრო მრავალფეროვან სცენარებს აცნობს და განზოგადების უნარს აუმჯობესებს[26].

[24] Mandeep Goyal and Qusay H. Mahmoud, “A Systematic Review of Synthetic Data Generation Techniques Using Generative AI,” *Electronics* 13, no. 17 (2024): 3509, https://doi.org/10.3390/electronics13173509.
[25] .Goyal and Mahmoud, “A Systematic Review of Synthetic Data Generation Techniques Using Generative AI.”
[26] Goyal and Mahmoud, “A Systematic Review of Synthetic Data Generation Techniques Using Generative AI.”

მონაცემების შექმნის სამი მთავარი მიდგომაა: გენერაციული მოწინააღმდეგე ქსელები (Generative Adversarial Networks - GANs), ვარიაციული ავტოენკოდერები (Variational Autoencoders - VAEs), დიფუზიური მოდელები (diffusion models) და დიდი ენობრივი მოდელები (Large Language Models - LLMs)[27].

GAN-ები შედგება ორი ნეირონული ქსელისგან. პირველი ნეირონული ქსელი - გენერატორი - ქმნის ხელოვნურ მონაცემს, ხოლო მეორე ქსელი - დისკრიმინატორი - მას აფასებს ნამდვილ მონაცემთან შედარებით. ქსელებს შორის უკუკავშირი აუმჯობესებს ხელოვნური მონაცემის ხარისხს, სანამ დეტექტორი ვერ განასხვავებს ხელოვნურ და ნამდვილ მონაცემს. VAE-ები შედგება ენკოდერისა და დეკოდერისგან. ენკოდერი ქმნის ფარულ სივრცეს (latent space - მონაცემთა მნიშვნელოვანი მახასიათებლების შემცველი სივრცე). ფარული სივრცის საშუალებით, დეკოდერი ქმნის ახალ მონაცემებს, რომლებიც ინარჩუნებს თავდაპირველი მონაცემების მთავარ თვისებებს[28]. დიფუზიური მოდელი წვრთნის დროს მონაცემებს უმატებს ახალ შემთხვევით კომპონენტებს (data noise). მაგალითად, თუ მოდელი იწვრთნება ფონდების ბირჟის მონაცემებზე, ის შემთხვევითობის პრინციპით ამატებს სხვადასხვა ფასის ფონდებს, სანამ საწვრთნელ მონაცემებში დამატებული შემთხვევითი კომპონენტები საწყისი ინფორმაციის დიდ ნაწილს არ ჩაანაცვლებს. შემდეგ ეტაპზე მოდელი ცდილობს, ნამდვილი მონაცემი შემთხვევით დამატებული მონაცემისგან გამოარჩიოს[29].

შეზღუდვები: GAN-ის ერთერთი ყველაზე დიდი პრობლემაა მოდების კოლაფსი (mode collapse) - როდესაც გენერატორი მონაცემთა განაწილების მხოლოდ გარკვეულ ტიპებს სწავლობს და საკმარისად მრავალფეროვან მონაცემებს ვერ ქმნის.

ასევე მნიშვნელოვანია, მკვლევარმა სწორი პარამეტრები გამოიყენოს, რათა გენერატორმა და დისკრიმინატორმა ერთმანეთი დააბალანსონ. ზედმეტად ძლიერი დისკრიმინატორი, რომელიც ყველა მონაცემს ყალბად თვლის, გენერატორს სწავლის საშუალებას არ აძლევს. VAE-ს პრობლემაა დაბალი ხარისხის მონაცემების შექმნა. ის მონაცემებს ყოველთვის სწორად ვერ აღადგენს, რადგან ფარული სივრციდან მცირე დეტალებს ტოვებს. შესაბამისად, მოდელს შეიძლება მაღალი განზომილების მონაცემთა გენერირება გაუჭირდეს. დიდ ენობრივ და

[27] Goyal and Mahmoud, “A Systematic Review of Synthetic Data Generation Techniques Using Generative AI.”
[28] Goyal and Mahmoud, “A Systematic Review of Synthetic Data Generation Techniques Using Generative AI.”
[29] Shuang Hao et al., “Synthetic Data in AI: Challenges, Applications, and Ethical Implications,” arXiv:2401.01629, preprint, arXiv, January 3, 2024, https://doi.org/10.48550/arXiv.2401.01629.

დიფუზიურ მოდელებს დიდი რესურსი სჭირდება - დამუშავების სიმძლავრე და მეხსიერება, შესაბამისად, მათი გამოყენება დიდ ხარჯებს ითვალისწინებს. დიფუზიური მოდელები ასევე დიდ დროს ანდომებს ხელოვნური მონაცემების გენერირებას და ბუნდოვანია, რომელ პრინციპს მიმართავს მოდელი მონაცემთა რეკონსტრუქციისას - როგორ აშორებს დამატებულ კომპონენტებს[30].

დიდი ენობრივი მოდელების გამოყენება მონაცემთა დამუშავებისთვის

ხშირ შემთხვევაში, ინფორმაციული ტექნოლოგიების უნარების ნაკლებობის გამო, კვლევებში პრობლემები იქმნება, მაგალითად, სოციალურ მეცნიერებებში, რომელი მიმართულებითაც მკვლევრებს დიდი კომპიუტერული მეცნიერებებისა და კოდირების ცოდნა არ გააჩნიათ. დიდი ენობრივი მოდელების საშუალებით შესაძლებელია კოდის გარეშე მონაცემების დამუშავება და მოდელების შექმნა.

მონაცემთა გაწმენდა და აღწერითი ანალიზი

პოპულარულ სტატისტიკურ კომპიუტერულ პროგრამებში ახლა ინტეგრირებულია სხვადასხვა ხელოვნური ინტელექტის ფუნქცია[31]. პროგრამები, როგორებიცაა პითონი (python) და R-ი, იყენებს ხელოვნური ინტელექტის ინსტრუმენტებს (მაგალითად Machlis, 2023), რაც მკვლევრებს კომპიუტერული პროგრამების ანალიტიკური შესაძლებლობების გაფართოების საშუალებას აძლევს. მაგალითად, მკვლევარი, რომელიც კომპლექსურ განგრძობით (longitudinal) კვლევაზე მუშაობს, შეუძლია ხელოვნური ინტელექტის გამოყენება მონაცემთა პირველადი დამუშავებისთვის და საწყისი მოდელების შერჩევაზე რეკომენდაციის მისაღებად. კერძოდ, მონაცემების გაწმენდა და დამუშავება ხელოვნური ინტელექტის საშუალებით შეიძლება შემდეგნაირად:

- ამოვარდნილი მაჩვენებლების დეტექციისთვის: unsupervised ML მოდელები (isolation forest, one-class SVM) უფრო ეფექტურად ამოიცნობენ მცდარ ან ამოვარდნილ

[30] Hafiz Muhammad Waseem et al., "Review of Generative AI for Synthetic Data Generation: A Healthcare Perspective," *Artificial Intelligence Review* 59, no. 2 (2025): 55, https://doi.org/10.1007/s10462-025-11440-2.
[31] Mike Perkins and Jasper Roe, "Generative AI Tools in Academic Research: Applications and Implications for Qualitative and Quantitative Research Methodologies," version 1, preprint, arXiv, 2024, https://doi.org/10.48550/ARXIV.2408.06872.

მაჩვენებლებს, ტრადიციულ ცალცვლადიან (univariate) სტატისტიკურ მოდელებთან შედარებით.

- დაკარგული მონაცემების ჩანაცვლება: ML მოდელებს, როგრიცაა K-nearest neighbours-ი, შეუძლიათ დაკარგული მონაცემის დაახლოებით დაანგარიშება. ეს ანგარიში უკეთესად მუშაობს, ვიდრე სტანდარტული მონაცემის საშუალო რიცხვით ან მედიანით ჩანაცვლება.
- ფონტების ინტეგრაცია: AutoML ინსტრუმენტისა და IBM Watson-ის საშუალებით შესაძლებელია სხვადასხვა ფორმატის (text, image, numerical) მონაცემების ანალიზი.
- მონაცემთა განზომილების შემცირება t-SNE-ით (t-distributed stochastic neighbor embedding) და UMAP-ით (uniform manifold approximation and projection).
- კატეგორიის პროგნოზირება მონაცემებზე დაფუძნებით: მაგალითად, ჩეთბოტის მიზანი იყო, განესაზღვრა, რომელი კატეგორია შეესაბამებოდა ქალს და რომელი - კაცს, ეს კატეგორიები მონაცემებში მხოლოდ 0-ისა და 1-ის სახით იყო კოდირებული. სხვა ცვლადებში არსებული ტრენდის მიხედვით, ჩეთბოტმა სწორი დასკვნა გააკეთა.

შემდეგ მკვლევარს შეუძლია მონაცემები ტრადიციული სტატისტიკური მეთოდებით გააანალიზოს და ინტერპრეტაციაში ხელოვნური ინტელექტი დაიხმაროს. ხელოვნურ ინტელექტს ასევე უნარი აქვს, რამდენიმე კომპიუტერულ პროგრამაში დააგენერიროს კოდი ერთი დავალებისთვის. ეს უზრუნველყოფს კოდის განზოგადებასა და თავსებადობას (generalization and compatibility)[32].

შეზღუდვები: ხელოვნური ინტელექტის მიერ გენერირებული კოდი კარგად მუშაობს მცირე დავალებებისთვის, მაგრამ მას უჭირს, თავი გაართვას კოდის გენერირებას უფრო კომპლექსური დავალებისთვის. ასევე, უძნელდება მუშაობა უცნობ დომენებში,[33] თუ მკვლევარი იყენებს არაოდრინალურ სტატისტიკურ მეთოდებს ან მონაცემთა ბაზას. ამ შემთხვევაში გენერირებული კოდი ხშირად შეიცავს შეცდომებს, ლოგიკურ ხარვეზებს, ან, ზოგადად, არაოპტიმალურია. ასევე ხელოვნური ინტელექტის მიერ გენერირებული კოდის გამოყენება ეთიკურ პრობლემებს უკავშირდება, განსაკუთრებით, მომხმარებლის მიერ შეყვანილი ინფორმაციის ვალიდაციის მხრივ.

[32] Perkins and Roe, “Generative AI Tools in Academic Research.”
[33] Hossein Hassani and Emmanuel Sirmal Silva, “The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field,” *Big Data and Cognitive Computing* 7, no. 2 (2023): 62, https://doi.org/10.3390/bdcc7020062.

პროგნოზირებადი ანალიტიკა

პროგნოზირებადი ანალიტიკის მიზანია მომავლის შეფასება ისტორიული მონაცემებისა და სტატისტიკური მოდელების მეშვეობით. ტრადიციული პროგნოზირებადი ანალიტიკა მოიხმობდა სტატისტიკურ მეთოდებს, როგორებიცაა რეგრესია ან დროითი მწკრივის ანალიზი. შესაბამის მოდელებს უჭირთ დიდი რაოდენობის არასტრუქტურულ მონაცემებთან გამკლავება[34]. რეგრესიის ძველი მოდელები ტიპურად ამუშავებდნენ 5-15 ცვლადს და მათი პროგნოზირების სიზუსტე, საშუალოდ, 60-70% იყო[35]. მოდელის პროგნოზირების შესაძლებობა უარსედება, როცა უფრო რთულ, მაღალგანზომილებიანი (high-dimensional)[36] მონაცემებისთვის იყენებენ[37].

ხელოვნური ინტელექტის განვითარებამ პროგნოზირებადი ანალიტიკის სფერო შეცვალა. გადაწყვეტილების ხეები (Decision Trees)[38], საყრდენი ვექტორების მანქანები (Support Vector Machines, SVM)[39] და ანსამბლური მეთოდები (Ensemble Methods)[40] მანქანური სწავლების მოდელებია, რომლებიც ანალიტიკისთვის გამოიყენება. ეს ალგორითმები უკეთესად უმკლავდება მრავალფეროვან მონაცემთა ბაზებს და უკეთესად ავლენს მონაცემთა შორის კომპლექსურ კავშირებს. პირველმა ხემ 15-20%-ით უფრო მაღალი პროგნოზირების სიზუსტე აჩვენა, ტრადიციულ სტატისტიკურ მოდელებთან შედარებით[41].

[34] Shubhodip Sasmal, "Predictive Analytics in Data Engineering: An AI Approach," *INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES* 12, no. 1 (2024): 13–18, https://doi.org/10.55083/irjeas.2024.v12i01003.

[35] Sasmal, "Predictive Analytics in Data Engineering."

[36] მონაცემები, რომლებშიც დიდი რაოდენობის ცვლადია. რაც უფრო დიდი რაოდენობის ცვლადია, მით უფრო მეტად უჭირს მოდელს გაარჩიოს, რომელი ცვლადებია მნიშვნელოვანი პროგნოზირებისთვის და რომლები - არა.

[37] Maheshkumar Mole, "Predictive Analytics Using AI: Forecasting the Future with Intelligent Insights," *Journal of Computer Science and Technology Studies* 7, no. 7 (2025): 214–19, https://doi.org/10.32996/jcsts.2025.7.7.20.

[38] მანქანური სწავლების ალგორითმი, რომელიც გამოიყენება კლასიფიკაციისა და რეგრესიისთვის. მას აქვს იერარქიული სტრუქტურა, შედგება "ტოტებისგან", "შიდა კვანძებისა" და "ფოთლებისგან". ალგორითმი ყველა ეტაპზე არჩევს კვანძებს (მონაცემს), რომლებიც ყველაზე ეფექტურად ყოფს მონაცემთა ბაზას. (წყარო - International Business Machines Corp. IBM.com)

[39] მანქანური სწავლების ალგორითმი, რომელიც გამოიყენება კლასიფიკაციისთვის. ალგორითმის მიზანია, იპოვოს საზღვარი (ჰიპერსიბრტყე), რომელიც კლასებს შორის დაშორებას მაქსიმალურად ზრდის და შესაბამისად, მონაცემებს ყველაზე ეფექტურად ანაწილებს. (წყარო - International Business Machines Corp. IBM.com)

[40] მანქანური სწავლების მოდელი, რომელიც უკეთესი პროგნოზირებისთვის რამდენიმე მოდელს აერთიანებს (მაგ. რეგრესიას და ხელოვნურ ნეირონულ ქსელს). (წყარო - International Business Machines Corp. IBM.com).

[41] Maheshkumar Mole, "Predictive Analytics Using AI."

ხელოვნური ინტელექტის სისტემებს შეუძლიათ რთულად ამოსაცნობი კანონზომიერების აღმოჩენა მაღალგანზომილებიან მონაცემებში. ნეირონული ქსელების (Convolutional Neural Network, Recurrent Neural Network) მეთოდების საშუალებით, შესაძლებელია დაუმუშავებელი მონაცემებიდან მნიშვნელოვანი მახასიათებლების გამოყოფა, რაც ტრადიციული მეთოდების გამოყენების პირობებში, მხოლოდ ცვლადების ხელით დამუშავებით იქნებოდა შესაძლებელი. ეს კი მნიშვნელოვან ანალიტიკურ რესურსს მოითხოვს.

ტრადიციული სტატისტიკური მოდელებისგან განსხვავებით, ხელოვნური ინტელექტის მოდელები მუდმივად სწავლობენ და ახალი მონაცემების მიღებისას, პროგნოზირებას აუმჯობესებენ. მაგალითად, სტრიმინგული მონაცემების დამუშავებისას, ადაპტირებადი მანქანური სწავლების მოდელების პროგნოზირების სიზუსტე 30 დღის განმავლობაში ყოველდღიურად 0.3-0.5%-ით იზრდებოდა. საბოლოოდ, მათ 22-34%-ით უფრო მაღალი პროგნოზირების სიზუსტე აჩვენეს, ვიდრე სტატიკურმა მოდელებმა[42].

შესაბამისი ანალიტიკისთვის გამოიყენება როგორც supervised learning-ის, ასევე unsupervised learning-ის ალგორითმები. Supervised learning-ი გულისხმობს ისეთი ალგორითმების გამოყენებას, რომლებიც ანოტირებული მონაცემების საშუალებით სწავლობენ. Gradient Boosting მეთოდები, როგორებიცაა XGBoost-ი, LightGBM-ი და CatBoost-ი, ხშირად გამოიყენება პროგნოზირებად ანალიტიკაში, რადგან სხვადასხვა ტიპის მონაცემების დამუშავება შეუძლია. Unsupervised learning-ი გულისხმობს ალგორითმებს, რომლებიც სწავლობენ არაანოტირებული მონაცემებით. DBSCAN-ი და K-means-ი ანალიტიკისთვის ყველაზე გამოყენებადი მეთოდებია[43].

შეზღუდვები: მეთოდებს ახლავს ხელოვნური ინტელექტისთვის დამახასიათებელი ტიპური შეზღუდვები. დიდი მნიშვნელობა აქვს იმ მონაცემთა ხარისხს, რომლითაც მოდელები სწავლობენ და შემდეგ პროგნოზს აკეთებენ. ასევე, ხელოვნური ინტელექტის მოდელების “black box” ალგორითმების გამო, შეცდომის დაშვების შემთხვევაში, რთულია იმის ამოცნობა, თუ სად არ გაამართლა მექანიზმმა, რაც მოდელის პრაქტიკულ გამოყენებას ზღუდავს.

[42] Maheshkumar Mole, “Predictive Analytics Using AI.”
[43] Sasmal, “Predictive Analytics in Data Engineering.”

ჯვარედინი ვალიდაცია

კლასიკური სხვაობათა სხვაობის (difference-in-difference) ან ინსტრუმენტული ცვლადების (instrumental variable (IV)) სტატისტიკური მეთოდების ნაცვლად, დღეს ხელოვნურ ინტელექტს შეუძლია აწარმოოს მოდელი, რომელიც ამცირებს განზოგადების ცდომილებას (generalization error)[44]. ერთ-ერთი ასეთი მეთოდია ჯვარედინი ვალიდაცია (cross validation): ის სტატისტიკურ მოდელებში გადაჭარბებული მორგების (overfitting)[45] პრევენციისთვის გამოიყენება. ჯვარედინი ვალიდაცია რამდენჯერმე ყოფს ერთსა და იმავე მონაცემების ბაზას - სატრენინგო და სატესტო ნაწილებად. მოდელი სატრენინგო მონაცემებს სავარჯიშოდ იყენებს, ამის შემდეგ შექმნილი მოდელი კი სატესტო ნაწილით მოწმდება. საბოლოო მოდელის შესაქმნელად, ეს პროცესი რამდენჯერმე მეორდება. მოდელის წარმატებულად შესაქმნელად საჭიროა საკმარისად დიდი მონაცემთა ბაზა, ხოლო სატესტო მონაცემები მთელი მონაცემთა ბაზის ნაკრებს ასახავდეს. პირველი ჯვარედინი ვალიდაციის მეთოდია მონაცემების დაყოფა სასწავლო და სატესტო ნაწილებად, მაგალითად, 70/30 პროპორციით. მეორე მიდგომაა K-1 ჯვარედინი ვალიდაცია, რომლის დროსაც მონაცემთა ნაკრები იყოფა K ნაწილად. მოდელის K-1 ნაწილი სატრენინგოა, ხოლო ბოლო ნაწილი მოდელის შესაფასებლად გამოიყენება. ეს პროცესი K-ჯერ მეორდება და საუკეთესო მოდელი loss ფუნქციით ვლინდება.

ვიზუალური ანალიტიკა

ხელოვნური ინტელექტის გამოყენებით შესაძლებელია უფრო სწრაფად დიდი რაოდენობის მონაცემების დამუშავება და ვიზუალიზაცია. სწორი ბრძანების (prompt) საშუალებით, შესაძლებელია ვიზუალიზაციების შექმნა.

დიდი ენობრივი მოდელების განვითარებამდე, მკვლევრები ვიზუალური ანალიტიკისთვის სხვა სისტემებს იყენებდნენ. ერთერთია V-NLIs-ი (Visualization-oriented Natural Language Interfaces), რომელიც ვიზუალიზაციებს მკვლევრის კონკრეტული დირექტივებით ქმნის. მაგალითად, თუ

[44] Jiaxing Zhang and Shuaishuai Feng, “Machine Learning Modeling: A New Way to Do Quantitative Research in Social Sciences in the Era of AI,” *Journal of Web Engineering*, ahead of print, March 16, 2021, https://doi.org/10.13052/jwe1540-9589.2023.

[45] გადაჭარბებული მორგება გულისხმობს მოდელის ზედმეტად მორგებას საწვრთნელ მონაცემებზე: შესაბამისად, მოდელი ანომალიებს ითვალისწინებს და ცუდი განზოგადების უნარი აქვს. ასეთი მოდელი მხოლოდ თავიდან მოცემულ საწვრთნელ მონაცემებზე აჩვენებს მაღალ პროგნოზირებით სიზუსტეს, რასაც ვერ ახერხებს ახალი მონაცემებით.

მკვლევარი მოითხოვს ცვლადების დროში ცვლილების ვიზუალიზაციას, აპლიკაცია ქმნის ხაზოვან დიაგრამას[46]. ასევე ხშირად გამოიყენება ვიზუალური ანალიტიკის რეკომენდაციის სისტემები, რომლებიც მკვლევარს სთავაზობს ვიზუალიზაციის რამდენიმე ვარიანტს, რომელთა შორის ის სასურველს შეარჩევს[47]. ამ სისტემებს ეფექტურად მუშაობისთვის სჭირდება დიდი რაოდენობის წესების პროგრამირება. თუ მკვლევარი კონკრეტულად იმ ფორმით არ მოითხოვს ვიზუალიზაციას, როგორი წესითაც სისტემაა პროგრამირებული, ის ვიზუალიზაციას ვერ შექმნის. ასეთი ტიპის შეზღუდვა არაა დამახასიათებელი დიდი ენობრივი მოდელებისთვის. წესების წინასწარ გაწერის ნაცვლად, მოდელებს ესმით ბუნებრივი ენა და შესაბამისად, შესაბამისი ვიზუალიზაციის შესაქმნელად მკვლევრის კონკრეტული დირექტივა არ სჭირდებათ.

ChatGPT-ის მიერ ორი ფუნქციის დამატებამ საგრძნობლად გაამარტივა მისი გამოყენება მონაცემთა დამუშავების პროცესში: ჩეთბოტში ფაილის ატვირთვის საშუალება და advanced data analysis მოდული/კოდის ინტერპრეტატორი[48]. კოდის ინტერპრეტატორი მონაცემთა ანალიზის და ვიზუალიზაციისთვის პითონს[49] იყენებს. ხელოვნური ინტელექტის დამატებითი ფუნქციების განვითარებასთან ერთად, უფრო და უფრო მარტივდება კარგი ვიზუალიზაციების შექმნა: მაგალითად, Claude-ს დაემატა ახალი ფუნქცია artifacts-ი, რომელიც აგენერირებს ვიზუალურ არტეფაქტებს, რაც მკვლევრებს კვლევის შედეგების წარდგენას უმარტივებს [50]. ასევე ერთ-ერთ კვლევაში წარმოდგენილია მონაცემთა ინტერპრეტატორი (Data Interpreter), LLM-ზე დაფუძნებული აგენტი, რომელიც შექმნილია რეალურ მონაცემებთან ადაპტაციის, ხელსაწყოების ინტეგრაციისა და მონაცემთა გენერაციის გამოწვევებთან გასამკლავებლად[51]. მონაცემთა ინტერპრეტატორი შეფასდა მონაცემთა ანალიზის სხვადასხვა ამოცანის საფუძველზე. მიღებულმა შედეგებმა აჩვენა, რომ მოდელმა აჯობა ღია კოდის საბაზისო მოდელებს, მათ შორის, ChatGPT-4-Turbo-ს[52].

[46] Maeve Hutchinson et al., “Foundation Model Assisted Visual Analytics: Opportunities and Challenges,” Computers & Graphics 130 (August 2025): 104246, https://doi.org/10.1016/j.cag.2025.104246.
[47] Hutchinson et al., “Foundation Model Assisted Visual Analytics.”
[48] Mike Perkins and Jasper Roe, “Generative AI Tools in Academic Research: Applications and Implications for Qualitative and Quantitative Research Methodologies,” version 1, preprint, arXiv, 2024, https://doi.org/10.48550/ARXIV.2408.06872.
[49] პითონი - პროგრამირების ენა
[50] Perkins and Roe, “Generative AI Tools in Academic Research.”
[51] Perkins and Roe, “Generative AI Tools in Academic Research.”
[52] Jie Min et al., “Applied Statistics in the Era of Artificial Intelligence: A Review and Vision,” arXiv:2412.10331, preprint, arXiv, December 13, 2024, https://doi.org/10.48550/arXiv.2412.10331.

***

ზემოთ განხილული კვლევების მიხედვით, ხელოვნური ინტელექტის ინტეგრაცია რაოდენობრივი კვლევის მეთოდოლოგიაში მკვლევრებს სამი ძირითადი მიმართულებით ეხმარება. პირველ რიგში, იგი ამცირებს მონაცემების შეგროვების, გაწმენდისა და დამუშავებისთვის საჭირო დროსა და რესურსებს. მეორე მხრივ, ხელოვნურ ინტელექტზე დაფუძნებული ინსტრუმენტები მონაცემთა ანალიზს უფრო ხელმისაწვდომს ხდის იმ მკვლევრებისთვის, რომლებსაც არ გააჩნიათ პროგრამირების ან მონაცემთა ანალიზის ღრმა ტექნიკური ცოდნა. მესამე, ხელოვნური ინტელექტის გამოყენება ხელს უწყობს გარკვეული ეთიკური გამოწვევების შემცირებას, მათ შორის, უზრუნველყოფს კონფიდენციალური ინფორმაციის დაცვას ისეთი მეთოდის გამოყენებით, როგორიცაა ხელოვნური მონაცემების გენერირება.

ამასთანავე, ხელოვნურ ინტელექტს მნიშვნელოვანი შეზღუდვებიც ახასიათებს. AI მოდელების ეფექტიანობა დიდწილად დამოკიდებულია იმ მონაცემების ხარისხზე, რომლებზეც ისინი გაწვრთნილია. შესაბამისად, არასრულყოფილმა ან მიკერძოებულმა სასწავლო მონაცემებმა შეიძლება არაზუსტი შედეგები აჩვენოს. გარდა ამისა, ხელოვნური ინტელექტის არაერთი მოდელი ე.წ. „შავი ყუთის“ (black box) პრინციპით ფუნქციონირებს, რის გამოც მკვლევრებისთვის ხშირად რთულია იმის გარკვევა, თუ როგორ მივიდა მოდელი კონკრეტულ გადაწყვეტილებამდე ან პროგნოზამდე. ამგვარმა გაუმჭვირვალობამ შეიძლება კვლევის შედეგების ინტერპრეტაცია და განმეორებადობა შეზღუდოს.