

ქართული ენის ციფრული ტრანსფორმაცია: ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების როლი

ირაკლი ბოჭორიშვილი¹; თინათინ კილაძე²; ქეთევან კვესელავა³; ედიშერ ჩქარეული⁴

¹საქართველოს ტექნიკური უნივერსიტეტის ასოცირებული პროფესორი, ინფორმატიკის დოქტორი, Orcid:0009-0006-4165-3149, Irakli.bochorishvili@gtu.ge; ²საქართველოს ტექნიკური უნივერსიტეტის პროფესორი, საერთაშორისო ურთიერთობების დოქტორი, https://orcid.org/0009-0009-9867-1294_tkiladze77@gmail.com; ³საქართველოს ტექნიკური უნივერსიტეტის პროფესორი, ინფორმატიკის ინჟინერიის აკადემიური დოქტორი, სსიპ დავით აღმაშენებლის სახელობის საქართველოს ეროვნული თავდაცვის აკადემიის სამეცნიერო-კვლევითი ცენტრის მთავარი მეცნიერი; https://orcid.org/0000-0002-1318-2750_q.veselava@gtu.ge; ⁴საქართველოს ტექნიკური უნივერსიტეტის ასოცირებული პროფესორი, მასობრივი კომუნიკაციის აკადემიური დოქტორი, <https://orcid.org/0009-0000-7753-0518>

აბსტრაქტი

ციფრული ტექნოლოგიების სწრაფმა განვითარებამ მნიშვნელოვნად შეცვალა თანამედროვე საზოგადოების კომუნიკაციური, საგანმანათლებლო და საინფორმაციო გარემო. განსაკუთრებული მნიშვნელობა შეიძინა ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების ინტეგრაციამ, რომელმაც შექმნა ახალი შესაძლებლობები, როგორც გლობალური, ისე ნაკლებად გავრცელებული ენებისათვის. ქართული ენა, როგორც უნიკალური დამწერლობისა და მრავალსაუკუნოვანი კულტურული მემკვიდრეობის მატარებელი ენა, თანამედროვე ციფრულ ეპოქაში მნიშვნელოვანი გამოწვევების წინაშე დგას. ერთის მხრივ, საჭიროა მისი სრულფასოვანი ინტეგრაცია თანამედროვე საინფორმაციო ტექნოლოგიებში, ხოლო მეორეს მხრივ — ენობრივი იდენტობისა და ნორმების დაცვა.

სტატიაში განხილულია ქართული ენის ციფრული ტრანსფორმაციის თანამედროვე ტენდენციები, ხელოვნური ინტელექტის როლი ენობრივი რესურსების განვითარებაში, ბუნებრივი ენის დამუშავების ტექნოლოგიების შესაძლებლობები, ავტომატური თარგმნის სისტემები, მეტყველების ამოცნობისა და სინთეზის პლატფორმები, მართლწერისა და გრამატიკის ავტომატური შემოწმების კორპუსების, ასევე დიდი ენობრივი მოდელების გავლენა ქართული ენის ციფრულ განვითარებაზე. განსაკუთრებული ყურადღება ეთმობა ქართული ენის კორპუსებს, ლექსიკური რესურსებისა და ციფრული ინფრასტრუქტურის განვითარების აუცილებლობას.

კვლევის მიზანია ქართული ენის ციფრული ტრანსფორმაციის ძირითადი მიმართულებების გაანალიზება, არსებული გამოწვევების შეფასება და იმ სტრატეგიული

ნაბიჯების განსაზღვრა, რომლებიც ხელს შეუწყობს ქართული ენის კონკურენტუნარიანობის ზრდას გლობალურ ციფრულ სივრცეში.

დასკვნები მიუთითებს, რომ ქართული ენის წარმატებული ციფრული ტრანსფორმაცია დიდწილად დამოკიდებულია ყოვლისმომცველი ენობრივი რესურსების, ღია მონაცემთა ინფრასტრუქტურის განვითარებაზე და აკადემიურ ინსტიტუტებს, სამთავრობო უწყებებსა და ტექნოლოგიურ სექტორს შორის მდგრად თანამშრომლობაზე. გარდა ამისა, ქართული ენის თანამედროვე ხელოვნური ინტელექტის სისტემებში ინტეგრაცია მოითხოვს ენის მოდელირების მიზანმიმართულ ძალისხმევას და გამოთვლითი ლინგვისტიკის კვლევაში ინვესტიციების გაზრდას.

სტატიაში ნათქვამია, რომ ქართული ენის ციფრული არსებობა არა მხოლოდ ტექნოლოგიური საკითხია, არამედ ხელოვნური ინტელექტის ეპოქაში კულტურული მდგრადობისა და ეროვნული სტრატეგიული განვითარების საკითხიც.

საკვანძო სიტყვები: ქართული ენა, ხელოვნური ინტელექტი, ენობრივი ტექნოლოგიები, ბუნებრივი ენის დამუშავება, ციფრული ტრანსფორმაცია, ენობრივი რესურსები, დიდი ენობრივი მოდელები.

შესავალი

XXI საუკუნე ციფრული ტრანსფორმაციის ეპოქად მიიჩნევა, როდესაც ინფორმაციული და საკომუნიკაციო ტექნოლოგიები პრაქტიკულად ყველა სფეროში იქცა განვითარების მთავარ მამოძრავებელ ძალად. ხელოვნური ინტელექტის სწრაფმა პროგრესმა არა მხოლოდ ტექნოლოგიური პროცესები შეცვალა, არამედ მნიშვნელოვანი გავლენა მოახდინა ენობრივ კომუნიკაციაზე, ცოდნის შექმნასა და გავრცელებაზე, განათლების სისტემასა და სახელმწიფო სერვისების ციფრულ განვითარებაზე. (Rishi Bommasani, 2021)

ბოლო ათწლეულის განმავლობაში განსაკუთრებული განვითარება განიცადა ბუნებრივი ენის დამუშავების ტექნოლოგიებმა, რომლებიც უზრუნველყოფენ ტექსტის ავტომატურ ანალიზს, კლასიფიკაციას, თარგმნას, შეჯამებას, დიალოგური სისტემების შექმნასა და ინფორმაციის ინტელექტუალურ დამუშავებას. დიდი ენობრივი მოდელების განვითარებამ მნიშვნელოვნად გაზარდა მრავალენოვანი ციფრული სისტემების შესაძლებლობები და შექმნა ახალი პერსპექტივები იმ ენებისთვისაც, რომელთა ციფრული რესურსები შეზღუდულია.

ქართული ენა მსოფლიო ენათა შორის ერთ-ერთი უნიკალური ენაა თავისი დამოუკიდებელი ანბანით, მდიდარი მორფოლოგიური სტრუქტურითა და მრავალსაუკუნოვანი წერილობითი ტრადიციით. ამასთანავე, მისი შედარებით მცირე ენობრივი ბაზა, შეზღუდული კორპუსები* და ნაკლებად განვითარებული ციფრული რესურსები ქმნიან გარკვეულ სირთულეებს თანამედროვე ხელოვნური ინტელექტის სისტემების განვითარებაში.

*კორპუსი არის ბუნებრივი მეტყველების ამსახველი მასალის ან ტექსტების მანქანურად წაკითხვადი კოლექცია, კორპუსი შესაძლოა შეიქმნას ენობრივი მასალის დეტალური შესწავლის მიზნით, ლინგვისტური თეორიების შესამუშავებლად, ან უკვე არსებული ლინგვისტური თეორიებისა თუ ჰიპოთეზების გადამოწმების მიზნით.

ციფრულ ეპოქაში ენის სიცოცხლისუნარიანობა აღარ განისაზღვრება მხოლოდ მისი ისტორიული მნიშვნელობით. თანამედროვე პირობებში გადამწყვეტი ხდება, თუ რამდენად ეფექტიანად არის ენა წარმოდგენილი ციფრულ გარემოში, რამდენად ხელმისაწვდომია მისი კორპუსები, ლექსიკონები, ანოტირებული მონაცემები, მეტყველების ბაზები და სხვა რესურსები, რომლებიც აუცილებელია ხელოვნური ინტელექტის მოდელების გასაწვდომად. (Pratik Joshi, 2020)

ქართული ენის ციფრული ტრანსფორმაცია მოიცავს რამდენიმე ურთიერთდაკავშირებულ მიმართულებას: ენობრივი მონაცემების ციფრულ ფორმატში დაგროვებას, ეროვნული კორპუსების შექმნას, ავტომატური თარგმნის სისტემების სრულყოფას, მეტყველების ამოცნობასა და სინთეზს, მართლწერისა და გრამატიკის ავტომატურ შემოწმებას, ინტელექტუალური საძიებო სისტემების განვითარებასა და დიალოგური ხელოვნური ინტელექტის ქართულენოვანი შესაძლებლობების გაფართოებას.

ამ პროცესს არა მხოლოდ ტექნოლოგიური, არამედ კულტურული, საგანმანათლებლო და ეროვნული მნიშვნელობაც გააჩნია. ენის ციფრული განვითარება პირდაპირ უკავშირდება ეროვნული კულტურული მემკვიდრეობის დაცვას, განათლების ხარისხის გაუმჯობესებას, სახელმწიფო ელექტრონული სერვისების ეფექტიანობასა და საერთაშორისო საინფორმაციო სივრცეში ქვეყნის კონკურენტუნარიანობას. (UNESCO, 2021; OECD, 2024).

აღსანიშნავია, რომ თანამედროვე მსოფლიოში ციფრული უთანასწორობა სულ უფრო მეტად განისაზღვრება არა მხოლოდ ინტერნეტზე წვდომით, არამედ იმითაც, თუ რამდენად სრულფასოვნად არის კონკრეტული ენა წარმოდგენილი ხელოვნური ინტელექტის სისტემებში. შესაბამისად, ქართული ენის ტექნოლოგიური განვითარება ეროვნული უსაფრთხოების, კულტურული დამოუკიდებლობისა და ციფრული სუვერენიტეტის მნიშვნელოვანი კომპონენტიც ხდება.

წინამდებარე კვლევის მიზანია ქართული ენის ციფრული ტრანსფორმაციის პროცესების კომპლექსური ანალიზი, ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების განვითარების ძირითადი მიმართულებების შეფასება, არსებული გამოწვევების იდენტიფიცირება და განვითარების პერსპექტიული რეკომენდაციების შემუშავება.

ასევე მიზანშეწონილად მივიჩნიეთ, რომ ჩაგვეტარებინა ემპირიული კვლევა, რომლის მიზანია, შეფასდეს ქართული ენის ციფრული ტრანსფორმაციის მიმდინარე მდგომარეობა, გამოიკვეთოს ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების გამოყენების პრაქტიკა, შეფასდეს არსებული პრობლემები და განისაზღვროს განვითარების ძირითადი მიმართულებები.

კვლევა ასევე მიზნად ისახავს დაადგინოს, რამდენად არის ქართული ენა ინტეგრირებული თანამედროვე ხელოვნური ინტელექტის სისტემებში, რამდენად იყენებენ სხვადასხვა პროფესიის წარმომადგენლები ქართულენოვან ციფრულ ინსტრუმენტებს და რა მოლოდინები აქვთ მათ აღნიშნული ტექნოლოგიების მიმართ.

კვლევის მეთოდოლოგია

კვლევა განხორციელდა რაოდენობრივი კვლევის მეთოდის გამოყენებით.

მონაცემების შეგროვება განხორციელდა სტრუქტურირებული კითხვარის მეშვეობით. კითხვარი მოიცავდა 5 კითხვას. (იხილეთ დანართი 1)

კვლევაში მონაწილეობა მიიღო 220 რესპონდენტმა, მათ შორის:

კატეგორია	რაოდენობა	%
სტუდენტები	82	37.3
უნივერსიტეტის აკადემიური პერსონალი	41	18.6
IT სპეციალისტები	36	16.4
საჯარო სექტორის თანამშრომლები	31	14.1
კერძო სექტორის წარმომადგენლები	30	13.6
	220	100

შედეგების შეჯამება

კვლევის შედეგები მიუთითებს რამდენიმე მნიშვნელოვან ტენდენციაზე.

პირველ რიგში, საქართველოში მნიშვნელოვნად იზრდება ხელოვნური ინტელექტის გამოყენების მასშტაბი, განსაკუთრებით განათლებისა და სამეცნიერო საქმიანობის სფეროებში. ეს მიუთითებს ციფრული ტრანსფორმაციის პროცესების გააქტიურებაზე და AI ტექნოლოგიების ყოველდღიურ პრაქტიკაში ინტეგრაციაზე.

მეორე მნიშვნელოვანი ტენდენცია უკავშირდება ქართული ენის მხარდაჭერის შეფასებას. მიუხედავად იმისა, რომ რესპონდენტები დადებითად აფასებენ ბოლო წლებში მიღწეულ პროგრესს, უმრავლესობა მიიჩნევს, რომ არსებული შესაძლებლობები ჯერ კიდევ არასაკმარისია. განსაკუთრებით პრობლემურ სფეროებად სახელდება სამეცნიერო ტერმინოლოგია, ავტომატური თარგმანი და მეტყველების ტექნოლოგიები.

კვლევის შედეგები ასევე ცხადყოფს, რომ რესპონდენტები პრიორიტეტულად მიიჩნევენ ეროვნული ენობრივი კორპუსების გაფართოებას, ღია მონაცემთა ბაზების შექმნასა და ქართული ენისთვის სპეციალიზებული AI მოდელების განვითარებას. ეს მიანიშნებს, რომ ტექნოლოგიური პროგრესი შეუძლებელია მაღალი ხარისხის ენობრივი რესურსების გარეშე.

ამგვარად ჩატარებულმა კვლევამ აჩვენა, რომ საქართველოში ხელოვნური ინტელექტის გამოყენება სწრაფად იზრდება და განსაკუთრებით აქტიურია აკადემიურ

და საგანმანათლებლო სივრცეში. ამავდროულად, ქართული ენის ტექნოლოგიური მხარდაჭერა კვლავ საჭიროებს სისტემურ განვითარებას. რესპონდენტთა შეფასებები ადასტურებს, რომ მომდევნო წლებში განსაკუთრებული ყურადღება უნდა დაეთმოს ეროვნული ტექსტური კორპუსების გაფართოებას, მეტყველების რესურსების შექმნას, ტერმინოლოგიური სტანდარტების განვითარებასა და უნივერსიტეტების, სახელმწიფო ინსტიტუტებისა და კერძო სექტორის თანამშრომლობის გაძლიერებას.

ძირითადი ნაწილი

ენობრივი ტექნოლოგიების განვითარება ბოლო სამი ათწლეულის განმავლობაში ერთ-ერთ ყველაზე სწრაფად მზარდ სამეცნიერო მიმართულებად ჩამოყალიბდა. ბუნებრივი ენის დამუშავების თანამედროვე მეთოდები ეფუძნება მანქანურ სწავლებას, ღრმა ნეირონულ ქსელებს, ტრანსფორმერულ არქიტექტურებსა და დიდი მოცულობის ტექსტურ კორპუსებს. (Brown 2020; OpenAI, 2023).

საერთაშორისო კვლევები აჩვენებს, რომ მაღალი ხარისხის ხელოვნური ინტელექტის სისტემების შექმნისათვის გადამწყვეტ მნიშვნელობას იმენს არა მხოლოდ ალგორითმების დახვეწა, არამედ მრავალფეროვანი, ხარისხიანი და ანოტირებული ენობრივი რესურსების არსებობა. (Bender & Friedman, 2022). სწორედ ამიტომ მსოფლიოს განვითარებული სახელმწიფოები მნიშვნელოვან ინვესტიციებს ახორციელებენ ეროვნული ენობრივი კორპუსების, ლექსიკური მონაცემთა ბაზებისა და მეტყველების რესურსების განვითარებაში.

ნაკლებად გავრცელებული ენებისთვის ერთ-ერთ ძირითად გამოწვევად რჩება მონაცემთა სიმცირე, დიალექტური მრავალფეროვნება, სტანდარტიზაციის საკითხები და ხარისხიანი ანოტირებული კორპუსების შეზღუდული მოცულობა. ეს ფაქტორები პირდაპირ აისახება ავტომატური თარგმნის, მეტყველების ამოცნობის, ტექსტის გენერაციისა და სხვა ენობრივი ტექნოლოგიების ეფექტიანობაზე.

ქართული ენის მიმართულებით ბოლო წლებში შეინიშნება მნიშვნელოვანი პროგრესი. შეიქმნა ახალი კორპუსები, ელექტრონული ლექსიკონები, მორფოლოგიური ანალიზატორები, მართლწერის შემმოწმებელი სისტემები და მანქანური თარგმნის ექსპერიმენტული მოდელები. (Kamarauli, 2024). მიუხედავად ამისა, არსებული რესურსები ჯერ კიდევ მნიშვნელოვნად ჩამორჩება იმ მოცულობასა და მრავალფეროვნებას, რომელიც აუცილებელია თანამედროვე დიდი ენობრივი მოდელების ეფექტიანი განვითარებისთვის. (Pratik Joshi et al., 2020).

ქართული ენის ციფრული ტრანსფორმაციის თეორიული საფუძვლები ციფრული ტრანსფორმაციის არსი და მნიშვნელობა

ციფრული ტრანსფორმაცია თანამედროვე საზოგადოების განვითარების ერთ-ერთ უმნიშვნელოვანეს პროცესს წარმოადგენს, რომელიც მოიცავს ინფორმაციული და საკომუნიკაციო ტექნოლოგიების სისტემურ ინტეგრაციას ეკონომიკაში, განათლებაში,

სახელმწიფო მმართველობაში, მეცნიერებასა და კულტურაში. აღნიშნული პროცესი მხოლოდ ტექნოლოგიური ინოვაციების დანერგვას არ გულისხმობს; იგი დაკავშირებულია ცოდნის შექმნის, დამუშავების, გავრცელებისა და გამოყენების ახალი მოდელების ჩამოყალიბებასთან. შესაბამისად, ციფრული ტრანსფორმაცია წარმოადგენს სოციალურ, ეკონომიკურ, კულტურულ და ტექნოლოგიურ ცვლილებათა ერთობლიობას, რომელიც მნიშვნელოვნად ცვლის საზოგადოების ფუნქციონირების ფორმებს. (European Commission, 2026).

ენის კონტექსტში ციფრული ტრანსფორმაცია გულისხმობს იმ პირობების შექმნას, რომლებიც უზრუნველყოფენ კონკრეტული ენის სრულფასოვან ფუნქციონირებას ციფრულ გარემოში. ეს მოიცავს ენობრივი მონაცემების ელექტრონულ ფორმატში წარმოდგენას, ტექსტური და ხმოვანი რესურსების ციფრულ დამუშავებას, ავტომატური ანალიზის სისტემების განვითარებას, ინტელექტუალური საძიებო მექანიზმების შექმნასა და ხელოვნური ინტელექტის პლატფორმებთან ინტეგრაციას. (UNESCO, 2021)

ციფრული ტრანსფორმაციის პირობებში ენის კონკურენტუნარიანობა მნიშვნელოვნად არის დამოკიდებული იმაზე, თუ რამდენად ფართოდაა წარმოდგენილი იგი თანამედროვე ტექნოლოგიურ სისტემებში. თუ ენა არ არის ინტეგრირებული ციფრულ პლატფორმებში, იგი თანდათან კარგავს ფუნქციურ შესაძლებლობებს, განსაკუთრებით განათლების, მეცნიერების, ბიზნესისა და საერთაშორისო კომუნიკაციის სფეროებში. ამიტომ ენის ციფრული განვითარება თანამედროვე სახელმწიფოს ერთ-ერთ სტრატეგიულ ამოცანად მიიჩნევა.

გლობალიზაციის პირობებში ციფრული ტრანსფორმაცია მნიშვნელოვან გავლენას ახდენს როგორც ფართოდ გავრცელებულ, ასევე მცირე ენებზე. საერთაშორისო კვლევები მიუთითებს, რომ იმ ენებს, რომელთაც გააჩნიათ განვითარებული ციფრული ინფრასტრუქტურა, ფართო კორპუსები და თანამედროვე ენობრივი ტექნოლოგიები, გაცილებით მეტი შესაძლებლობა აქვთ მონაწილეობა მიიღონ ხელოვნური ინტელექტის განვითარებაში. ამავდროულად, შეზღუდული რესურსების მქონე ენები დამატებითი გამოწვევების წინაშე დგანან, რადგან თანამედროვე ალგორითმების სწავლება მოითხოვს დიდ მოცულობის მონაცემებს.

ამ კონტექსტში ქართული ენის ციფრული ტრანსფორმაცია არ არის მხოლოდ ტექნოლოგიური განვითარების საკითხი. იგი მოიცავს ეროვნული კულტურული მემკვიდრეობის, ენობრივი იდენტობის, სამეცნიერო პოტენციალისა და ინფორმაციული დამოუკიდებლობის შენარჩუნებას. თანამედროვე პირობებში ენის განვითარება პირდაპირ უკავშირდება სახელმწიფოს ციფრულ სუვერენიტეტს, რადგან ხელოვნური ინტელექტის სისტემებში კონკრეტული ენის არარსებობა ან არასრულფასოვანი წარმოდგენა ზღუდავს მისი გამოყენების შესაძლებლობას მრავალ სფეროში.

ქართული ენის სპეციფიკა ციფრულ გარემოში

ქართული ენა მსოფლიო ენათა შორის გამოირჩევა უნიკალური ანბანით, მდიდარი მორფოლოგიური აგებულებითა და რთული გრამატიკული სისტემით. მისი აგლუტინაციური და ნაწილობრივ ფლექსიური სტრუქტურა, ზმნის მრავალფეროვანი

ფორმები, ბრუნვების სისტემა და სიტყვაწარმოების მექანიზმები, მნიშვნელოვან ტექნოლოგიურ გამოწვევებს ქმნიან ბუნებრივი ენის დამუშავების პროცესში. (Jurafsky & Martin, 2023).

განსაკუთრებული სირთულე უკავშირდება ზმნის ფორმათა მრავალფეროვნებას. ერთი ზმნის ლექსიკური ერთეულიდან შესაძლებელია ასეულობით განსხვავებული ფორმის წარმოქმნა, რაც მნიშვნელოვნად ართულებს მორფოლოგიურ ანალიზს, ავტომატურ ლემატიზაციას, სინტაქსურ ანოტაციასა და ტექსტის ავტომატურ დამუშავებას. ამასთანავე, ქართული ენის თავისუფალი სიტყვათა რიგი დამატებით ართულებს წინადადების სტრუქტურის ავტომატურ ანალიზს.

ქართული დამწერლობის უნიკალურობა ასევე განსაკუთრებულ მოთხოვნებს აყენებს პროგრამულ უზრუნველყოფასთან მიმართებით. მიუხედავად იმისა, რომ Unicode-ის სტანდარტის დანერგვამ მნიშვნელოვნად გააუმჯობესა ქართული ანბანის მხარდაჭერა, კვლავ არსებობს გამოწვევები ოპტიკური სიმბოლოების ამოცნობის (OCR), ისტორიული ტექსტების ციფრულიზაციის, ძველი ხელნაწერების ავტომატური დამუშავებისა და მრავალფონტიანი ტექსტების ანალიზის მიმართულებით. (Bender & Friedman, 2022).

თანამედროვე ციფრული ეკოსისტემა ეფუძნება დიდი მოცულობის ტექსტურ მონაცემებს. ქართულ ენაზე არსებული მონაცემები, მიუხედავად ბოლო წლებში მნიშვნელოვანი ზრდისა, კვლავ მნიშვნელოვნად ჩამორჩება ინგლისურ, გერმანულ, ფრანგულ, ესპანურ ან ჩინურ ენებზე არსებულ კორპუსებს. მონაცემთა შეზღუდული მოცულობა პირდაპირ აისახება ხელოვნური ინტელექტის მოდელების სიზუსტესა და შესაძლებლობებზე.

ენობრივი რესურსების მნიშვნელობა

ენობრივი რესურსი წარმოადგენს თანამედროვე ხელოვნური ინტელექტის საფუძველს. მათში მოიაზრება ტექსტური კორპუსები, ელექტრონული ლექსიკონები, ტერმინოლოგიური ბაზები, ანოტირებული მონაცემები, პარალელური ტექსტები, მეტყველების აუდიოკორპუსები, დიალექტური მონაცემები და სხვა ციფრული რესურსები.

თანამედროვე ნეირონული მოდელების განვითარება პირდაპირ არის დამოკიდებული მონაცემთა მოცულობასა და ხარისხზე. მაღალი ხარისხის ენობრივი რესურსები უზრუნველყოფს ტექსტის ავტომატური კლასიფიკაციის, ინფორმაციის ამოღების, მანქანური თარგმნის, მეტყველების ამოცნობის, ტექსტის გენერაციისა და დიალოგური სისტემების ეფექტიან მუშაობას.

ქართული ენისთვის განსაკუთრებული მნიშვნელობა ენიჭება ეროვნული კორპუსების გაფართოებას. აუცილებელია სხვადასხვა ჟანრის ტექსტების — მხატვრული ლიტერატურის, სამეცნიერო ნაშრომების, საკანონმდებლო დოკუმენტების, მედიის, სოციალურ ქსელებში გამოქვეყნებული მასალებისა და სასაუბრო მეტყველების — სისტემური შეგროვება და ანოტაცია. მონაცემთა მრავალფეროვნება ზრდის

ალგორითმების უნარს, უკეთ შეისწავლონ ენის სტრუქტურა და გამოყენების კონტექსტები.

ენობრივი რესურსების განვითარება მნიშვნელოვანია არა მხოლოდ ტექნოლოგიური პროგრესისთვის, არამედ ქართული კულტურული მემკვიდრეობის ციფრული შენარჩუნებისთვისაც. ძველი ხელნაწერების, ისტორიული დოკუმენტებისა და ლიტერატურული მემკვიდრეობის ციფრულიზაცია ქმნის შესაძლებლობას, ეს მასალა ხელმისაწვდომი გახდეს როგორც სამეცნიერო საზოგადოებისთვის, ასევე ხელოვნური ინტელექტის სისტემებისთვის.

ქართული ენის ციფრული ტრანსფორმაციის ძირითადი მიმართულებები

ქართული ენის თანამედროვე ციფრული განვითარება მოიცავს რამდენიმე სტრატეგიულ მიმართულებას:

პირველი მიმართულებაა ეროვნული ენობრივი მონაცემების სისტემური დაგროვება. აუცილებელია დიდი მოცულობის ღია ტექსტური და ხმოვანი რესურსების შექმნა, რომლებიც გამოყენებული იქნება კვლევით და ტექნოლოგიურ პროექტებში.

მეორე მიმართულებაა ბუნებრივი ენის დამუშავების ინსტრუმენტების განვითარება. ეს მოიცავს მორფოლოგიურ ანალიზატორებს,* სინტაქსურ პარსერებს,* სახელობითი ერთეულების ამომცნობ სისტემებს, ტექსტის კლასიფიკაციის მოდელებსა და სემანტიკური ანალიზის პლატფორმებს.

მესამე მიმართულებაა ავტომატური თარგმნის სისტემების გაუმჯობესება. ქართული ენის საერთაშორისო ინტეგრაციისთვის განსაკუთრებული მნიშვნელობა აქვს მრავალენოვანი ნეირონული მოდელების განვითარებას, რომლებიც უზრუნველყოფენ მაღალი ხარისხის თარგმანს ქართულსა და სხვა ენებს შორის.

მეოთხე მიმართულებაა მეტყველების ტექნოლოგიები, რომლებიც მოიცავს მეტყველების ამოცნობას, ხმოვანი ასისტენტების განვითარებასა და ტექსტის ხმოვან სინთეზს. აღნიშნული მიმართულება განსაკუთრებით მნიშვნელოვანია განათლების, სახელმწიფო სერვისებისა და ინკლუზიური ტექნოლოგიების განვითარებისათვის.

მეხუთე მიმართულებაა გენერაციული ხელოვნური ინტელექტის ქართულენოვანი შესაძლებლობების გაფართოება. თანამედროვე დიდი ენობრივი მოდელები უკვე მხარს უჭერენ ქართულ ენას, თუმცა მათი ხარისხი მნიშვნელოვანწილად დამოკიდებულია იმ მონაცემებზე, რომლებზეც ისინი ვითარდება. შესაბამისად, ქართული ენის ციფრული რესურსების ზრდა პირდაპირ განსაზღვრავს მომავალში ამ სისტემების ეფექტიანობასა და სიზუსტეს. (Vaswani, A, 2017); Vaswani et al., 2017; Brown et al., 2020).

*მორფოლოგიური ანალიზატორი არის პროგრამული ინსტრუმენტი, რომელიც ტექსტის სიტყვებს შლის კომპონენტებად (ფუძე, აფიქსი) და ადგენს მათ გრამატიკულ მახასიათებლებს (მეტყველების ნაწილი, ბრუნვა, პირი, დრო). მისი გამოყენება აადვილებს ტექსტის ავტომატიზებულ დამუშავებას, მთარგმნელობით პროგრამებსა და ორთოგრაფიული შეცდომების გასწორებას

*პარსერი (Parser) არის პროგრამული უზრუნველყოფა ან კოდის ნაწილი, რომელიც სხვადასხვა წყაროდან (ვებსაიტები, დოკუმენტები, ფაილები) არასტრუქტურირებულ ინფორმაციას კითხულობს, ამუშავებს და მომხმარებლისთვის ან პროგრამისთვის სასურველ, სტრუქტურირებულ ფორმატში (მაგალითად: Excel, JSON, CSV ან მონაცემთა ბაზა) გარდაქმნის

ხელოვნური ინტელექტი და ბუნებრივი ენის დამუშავების თანამედროვე ტექნოლოგიები

ხელოვნური ინტელექტის (Artificial Intelligence – AI) სწრაფმა განვითარებამ XXI საუკუნეში მნიშვნელოვნად შეცვალა ინფორმაციის შექმნის, დამუშავებისა და გავრცელების პროცესები. თუ თავდაპირველად ხელოვნური ინტელექტის სისტემები ეფუძნებოდა წესებზე დაფუძნებულ (Rule-Based) ალგორითმებს, თანამედროვე ეტაპზე დომინანტური გახდა მონაცემებზე დაფუძნებული მანქანური სწავლების (Machine Learning) და ღრმა სწავლების (Deep Learning) მეთოდები. ამ ცვლილებამ განსაკუთრებული გავლენა მოახდინა ბუნებრივი ენის დამუშავების (Natural Language Processing – NLP) განვითარებაზე. (Joshi et al., 2020).

ბუნებრივი ენის დამუშავება წარმოადგენს ინტერდისციპლინურ მიმართულებას, რომელიც აერთიანებს ლინგვისტიკას, კომპიუტერულ მეცნიერებას, მათემატიკასა და ხელოვნურ ინტელექტს. მისი ძირითადი მიზანია ისეთი ალგორითმების შექმნა, რომლებსაც შეუძლიათ ადამიანის ენის გააზრება, ანალიზი, გენერირება და ეფექტიანი გამოყენება. (Jurafsky & Martin, 2023). თანამედროვე NLP-ის ამოცანებია ტექსტის კლასიფიკაცია, ინფორმაციის ამოღება, ავტომატური თარგმნა, ემოციების ანალიზი, დიალოგური სისტემების შექმნა, ტექსტის შეჯამება, კითხვებზე პასუხის გაცემა და ტექსტის გენერაცია.

ბოლო ათწლეულის განმავლობაში განსაკუთრებული პროგრესი დაკავშირებულია ტრანსფორმერულ (Transformer) არქიტექტურასთან, რომელმაც შეცვალა ენობრივი მოდელების განვითარების მიდგომები. ტრანსფორმერული მოდელები მნიშვნელოვნად აჭარბებენ წინა თაობის რეკურენტულ და კონვოლუციურ ნეირონულ ქსელებს, რადგან შეუძლიათ გრძელ ტექსტებში კონტექსტური კავშირების ეფექტიანად დამუშავება. სწორედ ამ არქიტექტურაზეა დაფუძნებული თანამედროვე დიდი ენობრივი მოდელები (Large Language Models – LLMs), რომლებიც მრავალენოვან მონაცემებზე სწავლობენ და შეუძლიათ სხვადასხვა ენობრივი ამოცანის შესრულება. (Kamarauli, 2024).

თანამედროვე ბუნებრივი ენის დამუშავება მოიცავს რამდენიმე ძირითად მიმართულებას:

პირველი მიმართულებაა მორფოლოგიური ანალიზი, რომლის მიზანია სიტყვის გრამატიკული სტრუქტურის განსაზღვრა. ქართული ენისთვის ეს განსაკუთრებით მნიშვნელოვანია, რადგან სიტყვათა ფორმების მრავალფეროვნება, ზმნური სისტემის სირთულე და ბრუნვების სიმრავლე ავტომატურ ანალიზს მნიშვნელოვნად ართულებს.

მეორე მიმართულებაა სინტაქსური ანალიზი, რომელიც განსაზღვრავს სიტყვებს შორის გრამატიკულ კავშირებს. ქართული ენის თავისუფალი სიტყვათა რიგი ამ მიმართულებით დამატებით გამოწვევებს წარმოშობს, რის გამოც აუცილებელია სპეციალიზებული ალგორითმების განვითარება.

მესამე მიმართულებას წარმოადგენს სემანტიკური ანალიზი, რომლის ამოცანაა ტექსტის მნიშვნელობის ინტერპრეტაცია. თანამედროვე ნეირონული მოდელები

ცდილობენ არა მხოლოდ სიტყვების, არამედ მთლიანი ტექსტის კონტექსტური მნიშვნელობის გააზრებას.

მეოთხე მიმართულებაა ტექსტის გენერაცია, რომელიც გულისხმობს ადამიანის მსგავსი ტექსტების ავტომატურად შექმნას. ეს ტექნოლოგია გამოიყენება ჩატბოტებში, ინტელექტუალურ ასისტენტებში, საგანმანათლებლო პლატფორმებში, ჟურნალისტიკაში, პროგრამირებასა და სამეცნიერო საქმიანობაში.

განსაკუთრებული მნიშვნელობა ენიჭება მანქანურ თარგმანს, რომელიც თანამედროვე ნეირონული ქსელების გამოყენებით მნიშვნელოვნად გაუმჯობესდა. მიუხედავად ამისა, ქართული ენის სპეციფიკური გრამატიკული სტრუქტურა კვლავ წარმოადგენს მნიშვნელოვან გამოწვევას ავტომატური თარგმანის სისტემებისთვის.

დიდი ენობრივი მოდელები და ქართული ენა

დიდი ენობრივი მოდელები (Large Language Models – LLMs) ხელოვნური ინტელექტის განვითარების ერთ-ერთ ყველაზე მნიშვნელოვან მიღწევას წარმოადგენს. ისინი სწავლობენ მილიარდობით სიტყვაზე და ქმნიან ენის სტატისტიკურ წარმოდგენას, რაც მათ საშუალებას აძლევს შეასრულონ სხვადასხვა ამოცანა: ტექსტის შექმნა, თარგმნა, შეჯამება, კოდის გენერაცია, ინფორმაციის მოძიება და ანალიზი. (Stefanovitch et al., 2022).

ქართულ ენასთან მიმართებით დიდი ენობრივი მოდელების ეფექტიანობა მნიშვნელოვნად არის დამოკიდებული ქართულენოვანი მონაცემების რაოდენობასა და ხარისხზე. თუ მოდელი სწავლობს შეზღუდული მოცულობის ქართულ ტექსტებზე, მისი პასუხების სიზუსტე, ტერმინოლოგიური თანმიმდევრულობა და კონტექსტის გააზრების უნარი მცირდება.

ამ მიზეზით, ქართული ენისთვის განსაკუთრებული მნიშვნელობა ენიჭება ღია ტექსტური კორპუსების, აკადემიური ლიტერატურის, მედიის არქივების, საკანონმდებლო დოკუმენტების, ისტორიული ტექსტებისა და პარალელური კორპუსების სისტემურ გაფართოებას. რაც უფრო მრავალფეროვანია მონაცემები, მით უფრო მაღალი ხარისხის გახდება ქართულენოვანი ხელოვნური ინტელექტის სისტემები. (Mikaberidze et al., 2024).

თანამედროვე დიდი ენობრივი მოდელები მნიშვნელოვან როლს ასრულებენ განათლებაში, სახელმწიფო მმართველობაში, მეცნიერებაში, ბიზნესსა და მედიცინაში. მათ შეუძლიათ ავტომატურად შექმნან დოკუმენტები, შეაჯამონ დიდი მოცულობის ტექსტები, უპასუხოთ მომხმარებლის კითხვებს, დაეხმარონ პროგრამისტებს, მკვლევრებსა და სტუდენტებს. შესაბამისად, ქართული ენის სრულფასოვანი ინტეგრაცია ასეთ სისტემებში არა მხოლოდ ტექნოლოგიური, არამედ სტრატეგიული მნიშვნელობის ამოცანაა.

ხელოვნური ინტელექტის გამოყენების პრაქტიკული მიმართულებები ქართულ ენაში

ხელოვნური ინტელექტის ტექნოლოგიების პრაქტიკული დანერგვა ქართულენოვან გარემოში უკვე რამდენიმე მიმართულებით ვითარდება.

განათლების სფეროში AI გამოიყენება სასწავლო მასალების ავტომატური გენერაციისთვის, ტესტების შექმნისთვის, მართლწერისა და გრამატიკის შემოწმებისთვის, ინდივიდუალური სასწავლო რეკომენდაციების ფორმირებისა და სასწავლო პროცესის პერსონალიზაციისთვის.

სახელმწიფო სექტორში ხელოვნური ინტელექტი ხელს უწყობს დოკუმენტების კლასიფიკაციას, ადმინისტრაციული პროცესების ავტომატიზაციას, ელექტრონული სერვისების გაუმჯობესებასა და მოქალაქეთა მომსახურების ეფექტიანობის ზრდას.

სამეცნიერო კვლევებში AI გამოიყენება დიდი მოცულობის ტექსტური მონაცემების ანალიზისთვის, ბიბლიოგრაფიული ინფორმაციის ორგანიზებისთვის, სამეცნიერო ტექსტების შეჯამებისა და კვლევითი ტენდენციების იდენტიფიცირებისთვის.

მედიასა და ჟურნალისტიკაში ხელოვნური ინტელექტი გამოიყენება ახალი ამბების კლასიფიკაციისთვის, ავტომატური რედაქტირებისთვის, ტექსტების შეჯამებისა და კონტენტის გენერაციისთვის. თუმცა ამ პროცესში განსაკუთრებული ყურადღება უნდა დაეთმოს ინფორმაციის სიზუსტეს, წყაროების სანდოობასა და ეთიკურ სტანდარტებს.

ქართული ენის ციფრული განვითარების ძირითადი გამოწვევები

მიუხედავად მნიშვნელოვანი პროგრესისა, ქართული ენის ციფრული ტრანსფორმაცია კვლავ მრავალი გამოწვევის წინაშე დგას.

ერთ-ერთი მთავარი პრობლემაა მაღალხარისხიანი, ანოტირებული და ღია ენობრივი მონაცემების სიმცირე. თანამედროვე AI სისტემები ეფექტიანად ფუნქციონირებენ მხოლოდ მაშინ, როდესაც სწავლობენ მრავალფეროვან და დიდი მოცულობის მონაცემებზე. ქართულ ენაზე ასეთი რესურსები ჯერ კიდევ შეზღუდულია.

მეორე გამოწვევას წარმოადგენს ტერმინოლოგიური სტანდარტიზაცია. განსაკუთრებით ტექნოლოგიურ და სამეცნიერო სფეროებში ხშირად გამოიყენება უცხოენოვანი ტერმინები, რაც ართულებს როგორც ადამიანურ, ისე ავტომატურ კომუნიკაციას. აუცილებელია თანამედროვე ქართული სამეცნიერო და ტექნოლოგიური ტერმინოლოგიის სისტემური განვითარება.

მნიშვნელოვანი გამოწვევაა ასევე სპეციალისტების დეფიციტი. ქართული ენის ტექნოლოგიური განვითარება საჭიროებს ლინგვისტების, კომპიუტერული მეცნიერების სპეციალისტების, მონაცემთა ანალიტიკოსების, პროგრამისტებისა და ხელოვნური ინტელექტის მკვლევრების მჭიდრო თანამშრომლობას. სწორედ ინტერდისციპლინური თანამშრომლობა განსაზღვრავს ამ სფეროში წარმატების ხარისხს.

გარდა ამისა, აუცილებელია ეთიკური და სამართლებრივი ჩარჩოების განვითარება, რომლებიც უზრუნველყოფს პერსონალური მონაცემების დაცვას, საავტორო უფლებების პატივისცემას, ალგორითმების გამჭვირვალობასა და ხელოვნური ინტელექტის პასუხისმგებლიან გამოყენებას.

აღსანიშნავია, რომ ქართული ენის ციფრული ტრანსფორმაციის წარმატებისთვის საჭიროა სახელმწიფოს, აკადემიური და კერძო სექტორების კოორდინირებული თანამშრომლობა. სტრატეგიულ პრიორიტეტებად უნდა განისაზღვროს ეროვნული

ენობრივი კორპუსების გაფართოება, ღია მონაცემთა ბაზების შექმნა, დიალექტური და ისტორიული მასალის ციფრულიზაცია, მეტყველების რესურსების განვითარება, AI-ზე დაფუძნებული ქართულენოვანი პროგრამული უზრუნველყოფის მხარდაჭერა და საერთაშორისო კვლევით პროექტებში აქტიური მონაწილეობა.

განსაკუთრებული ყურადღება უნდა დაეთმოს უნივერსიტეტებსა და კვლევით ცენტრებში ენობრივი ტექნოლოგიების ლაბორატორიების განვითარებას, ახალგაზრდა მკვლევრების მომზადებასა და საერთაშორისო სამეცნიერო თანამშრომლობის გაძლიერებას. აღნიშნული ნაბიჯები შექმნის საფუძველს, რომ ქართული ენა სრულფასოვნად იყოს წარმოდგენილი თანამედროვე ციფრულ ეკოსისტემასა და ხელოვნური ინტელექტის მომავალ თაობის სისტემებში.

ქართული ენის ციფრული რესურსები და თანამედროვე ინფრასტრუქტურა

ხელოვნური ინტელექტისა და ბუნებრივი ენის დამუშავების სისტემების ეფექტიანობა პირდაპირ არის დამოკიდებული ენობრივი რესურსების ხარისხსა და მრავალფეროვნებაზე. თანამედროვე ალგორითმები სწავლობენ დიდი მოცულობის ტექსტურ და ხმოვან მონაცემებზე, რის საფუძველზეც ახერხებენ ენობრივი კანონზომიერებების ამოცნობას, კონტექსტის გააზრებასა და ახალი ტექსტების გენერირებას. შესაბამისად, ენობრივი რესურსები წარმოადგენს იმ ინფრასტრუქტურულ საფუძველს, რომელზეც შენდება თანამედროვე ციფრული ეკოსისტემა.

ენობრივი რესურსები მოიცავს ტექსტურ კორპუსებს, ელექტრონულ ლექსიკონებს, ტერმინოლოგიურ ბაზებს, მორფოლოგიურ და სინტაქსურ ანოტაციებს, პარალელურ კორპუსებს, მეტყველების აუდიომონაცემებს, დიალექტურ მასალებსა და ისტორიული ტექსტების ციფრულ არქივებს. თითოეულ მათგანს დამოუკიდებელი მნიშვნელობა აქვს, თუმცა მათი ერთობლივი გამოყენება ქმნის სრულფასოვან გარემოს თანამედროვე ხელოვნური ინტელექტის სისტემების განვითარებისათვის.

მსოფლიო პრაქტიკა ცხადყოფს, რომ იმ ქვეყნებს, რომლებმაც ეროვნული ენობრივი რესურსების შექმნა სახელმწიფო პოლიტიკის ნაწილად აქციეს, გაცილებით სწრაფად შეძლეს ხელოვნური ინტელექტის საკუთარ ენაზე განვითარება. სწორედ ამიტომ ქართული ენის ციფრული ინფრასტრუქტურის გაფართოება უნდა განიხილებოდეს როგორც ეროვნული მნიშვნელობის სტრატეგიული ამოცანა.

ქართულ ენაზე ბოლო წლებში შეიქმნა რამდენიმე მნიშვნელოვანი კორპუსი, რომლებმაც ხელი შეუწყო ლინგვისტიკური კვლევებისა და პროგრამული ინსტრუმენტების განვითარებას. მიუხედავად ამისა, მათი მოცულობა, ჟანრული მრავალფეროვნება და ანოტაციის ხარისხი ჯერ კიდევ ვერ პასუხობს თანამედროვე დიდი ენობრივი მოდელების მოთხოვნებს. (Kamarauli, M. 2024); (Irina Lobzhanidze, 2024); (ქეთევან მჭედლიშვილი, 2024).

განსაკუთრებული ყურადღება უნდა მიექცეს ისეთი ტექსტების ციფრულ დამუშავებას, როგორიცაა:

- კლასიკური და თანამედროვე ქართული ლიტერატურა;

- სამეცნიერო პუბლიკაციები;
- საკანონმდებლო და ადმინისტრაციული დოკუმენტები;
- მედიისა და საინფორმაციო სააგენტოების მასალები;
- აკადემიური სახელმძღვანელოები;
- დიალექტური ტექსტები;
- ისტორიული ხელნაწერები;
- სასაუბრო ენის ტრანსკრიფციები.

ასეთი მრავალფეროვანი მონაცემები მნიშვნელოვნად ზრდის ხელოვნური ინტელექტის უნარს, სწორად გაიაზროს ენის სტრუქტურა, ტერმინოლოგია და გამოყენების სხვადასხვა კონტექსტი.

ელექტრონული ლექსიკონები და ტერმინოლოგიური ბაზები

ელექტრონული ლექსიკონები თანამედროვე ენობრივი ტექნოლოგიების ერთ-ერთ ძირითად კომპონენტს წარმოადგენს. ისინი გამოიყენება როგორც ადამიანების მიერ ინფორმაციის მოსაძიებლად, ისე ავტომატური სისტემების მიერ სიტყვათა მნიშვნელობების, გრამატიკული მახასიათებლებისა და სემანტიკური კავშირების დასადგენად.

ქართული ენის შემთხვევაში განსაკუთრებული მნიშვნელობა ენიჭება მრავალდონიან ლექსიკოგრაფიულ რესურსებს, რომლებიც მოიცავს:

- განმარტებით ლექსიკონებს;
- ორთოგრაფიულ ლექსიკონებს;
- სინონიმებისა და ანტონიმების ბაზებს;
- ტერმინოლოგიურ ლექსიკონებს;
- ფრაზეოლოგიურ მონაცემთა ბაზებს;
- მრავალენოვან ლექსიკონებს.

სამეცნიერო და ტექნოლოგიური ტერმინოლოგიის სწრაფი განვითარება საჭიროებს მუდმივ განახლებას. განსაკუთრებით მნიშვნელოვანია ხელოვნური ინტელექტის, კიბერუსაფრთხოების, მონაცემთა მეცნიერების, ბიოინფორმატიკისა და ციფრული ტექნოლოგიების მიმართულებით ქართული ტერმინოლოგიის სტანდარტიზაცია. ერთიანი ტერმინოლოგიური პოლიტიკა ხელს შეუწყობს როგორც სამეცნიერო კომუნიკაციას, ისე ავტომატური თარგმნისა და ტექსტის ანალიზის სისტემების ხარისხის გაუმჯობესებას.

მეტყველების ტექნოლოგიები

ტექსტური მონაცემების გარდა, თანამედროვე ხელოვნური ინტელექტი ფართოდ იყენებს ხმოვან რესურსებს. მეტყველების ამოცნობის (Automatic Speech Recognition – ASR) და ტექსტის ხმოვან სინთეზის (Text-to-Speech – TTS) სისტემები სულ უფრო

მნიშვნელოვან როლს ასრულებს განათლებაში, მედიცინაში, სახელმწიფო სერვისებში, მობილურ ტექნოლოგიებსა და ინკლუზიურ კომუნიკაციაში.

ქართული ენისთვის ამ მიმართულებით ჯერ კიდევ არსებობს რამდენიმე მნიშვნელოვანი გამოწვევა:

- მაღალხარისხიანი აუდიოკორპუსების სიმცირე;
- რეგიონული დიალექტებისა და კილოების არასაკმარისი წარმოდგენა;
- ასაკობრივი და სოციალური მრავალფეროვნების ნაკლებობა;
- პროფესიული ტერმინოლოგიის შეზღუდული რაოდენობა;
- სპონტანური მეტყველების მცირე მოცულობა.

აღნიშნული პრობლემების გადასაჭრელად აუცილებელია მრავალსაათიანი ხმოვანი მონაცემების შეგროვება, ანოტაცია და ღია სამეცნიერო რესურსების სახით გამოქვეყნება. ეს მნიშვნელოვნად გააუმჯობესებს ქართულენოვანი ხმოვანი ასისტენტების, მობილური აპლიკაციებისა და სპეციალური საჭიროებების მქონე პირთათვის განკუთვნილი ტექნოლოგიების ხარისხს.

ქართული ენის ციფრულიზაცია და კულტურული მემკვიდრეობა

ქართული კულტურული მემკვიდრეობის მნიშვნელოვანი ნაწილი ჯერ კიდევ ხელნაწერების, არქივებისა და ბეჭდური გამოცემების სახით არის დაცული. მათი ციფრულიზაცია მხოლოდ ისტორიული დოკუმენტების შენარჩუნებას არ ემსახურება; იგი ქმნის ახალი თაობის კვლევით რესურსებს, რომლებიც შესაძლებელია გამოყენებულ იქნეს ხელოვნური ინტელექტის სისტემების სწავლებისთვის.

ისტორიული ტექსტების ციფრული დამუშავება რამდენიმე ეტაპს მოიცავს: დოკუმენტების სკანირებას, ოპტიკური სიმბოლოების ამოცნობას (OCR), ტექსტის კორექტირებას, ანოტაციასა და სტრუქტურირებულ მონაცემთა ბაზებში ინტეგრაციას. განსაკუთრებული მნიშვნელობა ენიჭება ძველი ქართული დამწერლობის (ასომთავრული, ნუსხური და მხედრული) ციფრულ მხარდაჭერას, რაც მომავალში შესაძლებელს გახდის ისტორიული ტექსტების ავტომატურ ანალიზსა და კვლევას.

კულტურული მემკვიდრეობის ციფრულიზაცია ასევე ხელს უწყობს საერთაშორისო სამეცნიერო თანამშრომლობას, რადგან ქართული ისტორიული წყაროები უფრო ხელმისაწვდომი ხდება უცხოელი მკვლევარებისათვის.

მიუხედავად ბოლო წლებში განხორციელებული მნიშვნელოვანი პროექტებისა, ქართული ენის ციფრული ინფრასტრუქტურა ჯერ კიდევ განვითარების ეტაპზეა. არსებული გამოწვევები რამდენიმე ძირითად ჯგუფად შეიძლება დაიყოს.

პირველი გამოწვევა დაკავშირებულია მონაცემთა მოცულობასთან. თანამედროვე გენერაციული ხელოვნური ინტელექტის სისტემების სწავლებისთვის საჭიროა მილიარდობით სიტყვა, მაშინ როდესაც ქართულ ენაზე ღია და სტრუქტურირებული მონაცემები შედარებით შეზღუდულია.

მეორე პრობლემა არის რესურსების ფრაგმენტაცია. სხვადასხვა ინსტიტუციის მიერ შექმნილი კორპუსები, ლექსიკონები და მონაცემთა ბაზები ხშირად ერთმანეთთან

შეუთავსებელია, რაც ზღუდავს მათი ერთიან ეკოსისტემაში ინტეგრაციის შესაძლებლობას.

მესამე გამოწვევაა სამართლებრივი და ეთიკური საკითხები. აუცილებელია საავტორო უფლებების, პერსონალური მონაცემების დაცვისა და ღია ლიცენზიების ისეთი სისტემის განვითარება, რომელიც ერთდროულად დაიცავს ავტორების ინტერესებს და ხელს შეუწყობს სამეცნიერო კვლევებს.

მეოთხე გამოწვევას წარმოადგენს კვალიფიციური ადამიანური რესურსის სიმცირე. თანამედროვე ენობრივი ტექნოლოგიების განვითარება მოითხოვს ლინგვისტიკის, პროგრამისტების, მონაცემთა მეცნიერების სპეციალისტების, ხელოვნური ინტელექტის მკვლევრებისა და დარგობრივი ექსპერტების ინტერდისციპლინურ თანამშრომლობას.

აღსანიშნავია, რომ ქართული ენის ციფრული განვითარების გრძელვადიანი სტრატეგია უნდა ეფუძნებოდეს რამდენიმე ძირითად პრინციპს:

- ეროვნული ენობრივი რესურსების სისტემური გაფართოება;
- ღია მონაცემთა პოლიტიკის მხარდაჭერა;
- აკადემიური, სახელმწიფო და კერძო სექტორების თანამშრომლობის გაძლიერება;
- ქართული ენის ტექნოლოგიური სტანდარტების შემუშავება;
- საერთაშორისო კვლევით პროექტებში ინტეგრაცია;
- ხელოვნური ინტელექტის კვლევისა და განვითარების ეროვნული ცენტრების მხარდაჭერა;
- ახალგაზრდა მკვლევრებისა და სპეციალისტების მომზადება.

აღნიშნული მიმართულებების თანმიმდევრული განხორციელება უზრუნველყოფს არა მხოლოდ ქართული ენის სრულფასოვან ინტეგრაციას თანამედროვე ციფრულ ეკოსისტემაში, არამედ გაზრდის საქართველოს სამეცნიერო, საგანმანათლებლო და ტექნოლოგიურ კონკურენტუნარიანობას საერთაშორისო დონეზე.

საერთაშორისო გამოცდილება მცირე ენების ციფრული ტრანსფორმაციის სფეროში მცირე ენების ციფრული განვითარების თანამედროვე ტენდენციები

ციფრული ტრანსფორმაციის ეპოქაში ენის სიცოცხლისუნარიანობა სულ უფრო მეტად განისაზღვრება მისი წარმომადგენლობით ციფრულ გარემოში. თუ წარსულში ენის განვითარების ძირითად ფაქტორებად მიიჩნეოდა დემოგრაფიული მაჩვენებლები, განათლების სისტემა და კულტურული ტრადიცია, დღეს გადამწყვეტი მნიშვნელობა ენიჭება იმას, თუ რამდენად არის კონკრეტული ენა წარმოდგენილი ხელოვნური ინტელექტის სისტემებში, ციფრულ პლატფორმებსა და ენობრივ ტექნოლოგიებში.

მცირე ენებისათვის ეს გამოწვევა განსაკუთრებით მნიშვნელოვანია, რადგან თანამედროვე გენერაციული ხელოვნური ინტელექტის სისტემები ეფექტიანად ფუნქციონირებენ მხოლოდ იმ შემთხვევაში, როდესაც მათ ხელმისაწვდომი აქვთ დიდი მოცულობის, მრავალფეროვანი და მაღალხარისხიანი ენობრივი მონაცემები. სწორედ

ამიტომ არაერთმა ქვეყანამ ეროვნული ენობრივი რესურსების განვითარება სახელმწიფო სტრატეგიის ნაწილად აქცია. (OECD, 2024).

ესტონეთის გამოცდილება

ესტონეთი ერთ-ერთ ყველაზე წარმატებულ მაგალითად მიიჩნევა მცირე ენის ციფრული განვითარების მიმართულებით. ქვეყნის ციფრული განვითარების პოლიტიკა ეფუძნება სახელმწიფოს, უნივერსიტეტებისა და კერძო სექტორის მჭიდრო თანამშრომლობას. განსაკუთრებული ყურადღება ეთმობა ეროვნული ენობრივი კორპუსების განვითარებას, სახელმწიფო დოკუმენტების ციფრულ ფორმატში ხელმისაწვდომობას, ელექტრონულ მმართველობასა და ხელოვნური ინტელექტის დანერგვას საჯარო სერვისებში.

ესტონეთის გამოცდილება ცხადყოფს, რომ მცირე მოსახლეობის მიუხედავად შესაძლებელია ისეთი ენობრივი ინფრასტრუქტურის შექმნა, რომელიც უზრუნველყოფს ეროვნული ენის ეფექტიან გამოყენებას როგორც ადმინისტრაციულ, ისე საგანმანათლებლო და სამეცნიერო სივრცეში. (European Commission, 2026).

ფინეთის გამოცდილება

ფინეთი განსაკუთრებულ ყურადღებას უთმობს ღია მონაცემების პოლიტიკას. სახელმწიფო ინსტიტუტები, უნივერსიტეტები და კვლევითი ცენტრები სისტემატურად ქმნიან და აქვეყნებენ ღია ლიცენზიით ხელმისაწვდომ ტექსტურ და ხმოვან რესურსებს, რაც მნიშვნელოვნად უწყობს ხელს ენობრივი ტექნოლოგიების განვითარებას.

ფინეთის გამოცდილება აჩვენებს, რომ ღია მონაცემთა ბაზები არა მხოლოდ სამეცნიერო კვლევების განვითარებას უწყობს ხელს, არამედ აჩქარებს ინოვაციური პროდუქტებისა და მომსახურებების შექმნას კერძო სექტორშიც. (UNESCO, 2021).

ისლანდიის გამოცდილება

ისლანდია წარმოადგენს წარმატებულ მაგალითს იმისა, თუ როგორ შეიძლება მცირე მოსახლეობის მქონე სახელმწიფომ დაიცვას საკუთარი ენა გლობალური ტექნოლოგიური ცვლილებების პირობებში.

ისლანდიის სახელმწიფო მრავალი წლის განმავლობაში ახორციელებს სპეციალურ პროგრამებს, რომლებიც მიზნად ისახავს მეტყველების რესურსების, ეროვნული კორპუსების, ლექსიკური მონაცემთა ბაზებისა და ხელოვნური ინტელექტისთვის საჭირო ენობრივი მონაცემების განვითარებას. განსაკუთრებული ყურადღება ეთმობა თანამედროვე ტექნოლოგიებში ისლანდიური ენის სრულფასოვან გამოყენებას, რათა მოსახლეობას არ მოუწიოს უცხო ენაზე გადასვლა ციფრული სერვისებით სარგებლობისას. (UNESCO, 2021).

უელსური და ბასკური ენების გამოცდილება

უელსური ენის ციფრული განვითარება წარმოადგენს მცირე ენების ტექნოლოგიური მხარდაჭერის ერთ-ერთ წარმატებულ მაგალითს. ბოლო წლებში მნიშვნელოვნად გაიზარდა უელსური ენის ინტეგრაცია ციფრულ პლატფორმებსა და ხელოვნური

ინტელექტის სისტემებში. სახელმწიფოს, აკადემიური ინსტიტუტებისა და ტექნოლოგიური კომპანიების თანამშრომლობით შეიქმნა ენობრივი კორპუსები, მეტყველების ამოცნობის სისტემები, ავტომატური თარგმნის ინსტრუმენტები და სხვა ენობრივი რესურსები, რომლებმაც მნიშვნელოვნად გააუმჯობესა უელსური ენის გამოყენება ციფრულ გარემოში (Jones, 2022; UNESCO, 2021).

განსაკუთრებული მნიშვნელობა ენიჭება ღია ენობრივი რესურსების შექმნასა და მათ ხელმისაწვდომობის მკვლევარებისა და პროგრამული უზრუნველყოფის შემქმნელებისთვის. აღნიშნულმა მიდგომამ ხელი შეუწყო უელსური ენის ინტეგრაციას თანამედროვე ბუნებრივი ენის დამუშავების (NLP) სისტემებს, მრავალენოვან ხელოვნური ინტელექტის მოდელებში. (European Commission, 2026; OECD, 2023).

უელსის გამოცდილება ცხადყოფს, რომ მცირე ენის ტექნოლოგიური განვითარება შესაძლებელია მხოლოდ გრძელვადიანი სახელმწიფო პოლიტიკის, აკადემიური თანამშრომლობისა და ციფრული ინფრასტრუქტურის სისტემური განვითარების პირობებში (UNESCO, 2021).

ბასკური ენა ერთ-ერთ ყველაზე წარმატებულ მაგალითად მიიჩნევა ევროპაში ენობრივი ტექნოლოგიების განვითარების მიმართულებით. მნიშვნელოვანი როლი ამ პროცესში ეკისრება აკადემიურ და კვლევით ცენტრებს, რომლებმაც შექმნეს ფართომასშტაბიანი ტექსტური კორპუსები, ლექსიკური ბაზები, მორფოლოგიური ანალიზატორები, სინტაქსური ანალიზატორები და ავტომატური თარგმნის სისტემები (Agerri et al., 2020).

ბასკური ენის განვითარების ერთ-ერთი მთავარი ფაქტორია ღია მონაცემების პოლიტიკა და სამეცნიერო რესურსების ფართო ხელმისაწვდომობა, რაც უზრუნველყოფს ახალი AI მოდელების ეფექტიან სწავლებას და მუდმივ გაუმჯობესებას (European Commission, 2026).

კვლევები მიუთითებს, რომ ბასკური ენის ტექნოლოგიური წარმატება ეფუძნება უნივერსიტეტების, კვლევითი ინსტიტუტებისა და სახელმწიფოს კოორდინირებულ თანამშრომლობას, რაც მნიშვნელოვნად ზრდის მცირე ენის კონკურენტუნარიანობას გლობალურ ციფრულ გარემოში (Joshi et al., 2020; OECD, 2024).

ბასკური გამოცდილება საქართველოსთვის განსაკუთრებით საინტერესოა, რადგან ორივე ენა გამოირჩევა უნიკალური გრამატიკული სტრუქტურით და შეზღუდული რაოდენობის ციფრული რესურსებით. შესაბამისად, ბასკური ენის ტექნოლოგიური განვითარების მოდელი შეიძლება გამოყენებულ იქნეს ქართული ენის ციფრული ტრანსფორმაციის სტრატეგიის შემუშავებისას (UNESCO, 2021).

უელსური ენისა და ბასკური ენის განვითარების პროგრამები აჩვენებს, რომ ციფრული ტექნოლოგიები მნიშვნელოვან როლს ასრულებენ ენის შენარჩუნებასა და პოპულარიზაციაში.

ამ ქვეყნების გამოცდილებაში განსაკუთრებული ადგილი უკავია:

- სასკოლო განათლების ციფრულ რესურსებს;
- ხმოვანი ტექნოლოგიების განვითარებას;

- მრავალენოვანი ელექტრონული ლექსიკონების შექმნას;
- მობილური აპლიკაციების განვითარებას;
- ახალგაზრდებში ეროვნული ენის პოპულარიზაციას სოციალური მედიისა და ციფრული პლატფორმების მეშვეობით.

ეს გამოცდილება ადასტურებს, რომ ენის დაცვა თანამედროვე ეპოქაში აღარ შემოიფარგლება მხოლოდ საგანმანათლებლო პოლიტიკით და მოითხოვს თანამედროვე ტექნოლოგიების სისტემურ გამოყენებას.

საერთაშორისო გამოცდილების მნიშვნელობა საქართველოსთვის

საერთაშორისო პრაქტიკის ანალიზი ცხადყოფს, რომ მცირე ენების წარმატებული ციფრული ტრანსფორმაცია ეფუძნება რამდენიმე საერთო პრინციპს.

პირველ რიგში, აუცილებელია ეროვნული ენობრივი რესურსების განვითარების გრძელვადიანი სახელმწიფო პოლიტიკა. მეორე, განსაკუთრებული მნიშვნელობა აქვს ღია მონაცემთა ბაზების შექმნას, რაც ხელს უწყობს როგორც აკადემიურ კვლევებს, ასევე კერძო სექტორის ინიციატივებს. მესამე, მნიშვნელოვანია უნივერსიტეტების, კვლევითი ინსტიტუტების, ტექნოლოგიური კომპანიებისა და საჯარო სექტორის თანამშრომლობა. მეოთხე, საჭიროა საერთაშორისო პროექტებში აქტიური მონაწილეობა, რაც უზრუნველყოფს ცოდნის, გამოცდილებისა და ტექნოლოგიური რესურსების გაზიარებას.

საქართველოსთვის აღნიშნული გამოცდილება განსაკუთრებით მნიშვნელოვანია, რადგან ქართული ენა, როგორც უნიკალური დამწერლობისა და მდიდარი კულტურული მემკვიდრეობის მქონე ენა, საჭიროებს მიზანმიმართულ და გრძელვადიან პოლიტიკას, რომელიც უზრუნველყოფს მის სრულფასოვან ფუნქციონირებას ხელოვნური ინტელექტისა და ციფრული ტექნოლოგიების ეპოქაში.

ქართული ენის ციფრული ტრანსფორმაციის განვითარების სტრატეგიული რეკომენდაციები

➤ ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების სწრაფი განვითარება ცვლის თანამედროვე საზოგადოების ფუნქციონირების მოდელს. ამ პირობებში ქართული ენის ციფრული განვითარება აღარ წარმოადგენს მხოლოდ ლინგვისტურ ან ტექნოლოგიურ საკითხს. იგი დაკავშირებულია ქვეყნის სამეცნიერო, საგანმანათლებლო, ეკონომიკურ, კულტურულ და საინფორმაციო უსაფრთხოებასთან. შესაბამისად, საჭიროა ერთიანი ეროვნული სტრატეგიის ჩამოყალიბება, რომელიც უზრუნველყოფს ქართული ენის სრულფასოვან ფუნქციონირებას თანამედროვე ციფრულ ეკოსისტემაში.

ასეთი სტრატეგია უნდა ეფუძნებოდეს სახელმწიფოს, აკადემიური სექტორის, კერძო კომპანიებისა და საერთაშორისო პარტნიორების თანამშრომლობას. მხოლოდ კოორდინირებული ძალისხმევით იქნება შესაძლებელი ისეთი ინფრასტრუქტურის შექმნა, რომელიც ქართული ენის ტექნოლოგიურ განვითარებას გრძელვადიან საფუძველს შეუქმნის.

- სახელმწიფომ უნდა განიხილოს ქართული ენის ციფრული განვითარება როგორც ეროვნული მნიშვნელობის პროექტი. ამ მიზნით მიზანშეწონილია:

პირველ რიგში, შემუშავდეს ქართული ენის ციფრული განვითარების ეროვნული სტრატეგია, რომელშიც მკაფიოდ განისაზღვრება პრიორიტეტები, პასუხისმგებელი უწყებები, დაფინანსების მექანიზმები და შესრულების ინდიკატორები.

აუცილებელია ეროვნული ენობრივი რესურსების — ტექსტური კორპუსების, მეტყველების მონაცემთა ბაზების, დიალექტური მასალებისა და ტერმინოლოგიური რესურსების — შექმნისა და განახლების სისტემური მხარდაჭერა. მნიშვნელოვანია, რომ აღნიშნული რესურსები, სამართლებრივი და ეთიკური მოთხოვნების დაცვით, მაქსიმალურად ხელმისაწვდომი იყოს სამეცნიერო და ტექნოლოგიური კვლევებისთვის.

სახელმწიფო პოლიტიკის მნიშვნელოვან მიმართულებად უნდა იქცეს ღია მონაცემების (Open Data) ინიციატივების მხარდაჭერა. ღია მონაცემები ხელს უწყობს როგორც აკადემიური კვლევის, ისე ინოვაციური ტექნოლოგიური პროდუქტების განვითარებას.

ასევე მნიშვნელოვანია, რომ ელექტრონული მმართველობის სისტემებში ხელოვნური ინტელექტის დანერგვისას უზრუნველყოფილი იყოს ქართული ენის სრულფასოვანი მხარდაჭერა, რათა მოქალაქეებს სახელმწიფო სერვისებზე წვდომა მშობლიურ ენაზე ჰქონდეთ.
- უნივერსიტეტები უნდა იქცნენ ქართული ენის ტექნოლოგიური განვითარების ერთ-ერთ მთავარ ცენტრად. მიზანშეწონილია ინტერდისციპლინური სასწავლო პროგრამების განვითარება, რომლებიც გააერთიანებს ლინგვისტიკას, კომპიუტერულ მეცნიერებას, მონაცემთა ანალიზსა და ხელოვნურ ინტელექტს.
- განსაკუთრებული ყურადღება უნდა დაეთმოს კვლევითი ლაბორატორიების შექმნას, სადაც განხორციელდება ქართული ენის კორპუსების ანოტაცია, მორფოლოგიური და სინტაქსური ანალიზის ინსტრუმენტების განვითარება, მეტყველების ტექნოლოგიების კვლევა და გენერაციული ხელოვნური ინტელექტის ქართულენოვანი მოდელების შეფასება.
- მნიშვნელოვანია სტუდენტების ჩართვა პრაქტიკულ კვლევით პროექტებში, რადგან ეს ხელს შეუწყობს ახალი თაობის სპეციალისტების მომზადებას, რომლებიც შეძლებენ ენობრივი ტექნოლოგიების განვითარებას როგორც საქართველოში, ისე საერთაშორისო სამეცნიერო სივრცეში.
- სამეცნიერო კვლევების განვითარება უნდა ეფუძნებოდეს ინტერდისციპლინურ თანამშრომლობას. თანამედროვე ენობრივი ტექნოლოგიების შექმნა შეუძლებელია მხოლოდ ლინგვისტების ან მხოლოდ პროგრამისტების ძალისხმევით. აუცილებელია ერთობლივი მუშაობა ლინგვისტების, ხელოვნური ინტელექტის მკვლევრების, მონაცემთა მეცნიერების სპეციალისტების, მათემატიკოსებისა და დარგობრივი ექსპერტების მონაწილეობით.

განსაკუთრებული მნიშვნელობა აქვს კვლევითი შედეგების ღია გამოქვეყნებას, რეპროდუცირებადობასა და საერთაშორისო სამეცნიერო თანამშრომლობას. ქართულ ენაზე შექმნილი რესურსები უნდა აკმაყოფილებდეს საერთაშორისო სტანდარტებს, რაც გაზრდის მათ გამოყენებადობას გლობალურ კვლევით პროექტებში.

- კერძო სექტორს შეუძლია მნიშვნელოვანი წვლილი შეიტანოს ქართული ენის ციფრულ განვითარებაში. სასურველია, ტექნოლოგიურმა კომპანიებმა აქტიურად გამოიყენონ ქართულენოვანი ინტერფეისები, განავითარონ ხმოვანი ასისტენტები, ჩატბოტები, ავტომატური თარგმნის სისტემები და მართლწერისა და გრამატიკის შემოწმების ინსტრუმენტები.

ასევე მიზანშეწონილია თანამშრომლობა უნივერსიტეტებთან და კვლევით ცენტრებთან, რათა სამეცნიერო მიღწევები უფრო სწრაფად გადაიზარდოს პრაქტიკულ ტექნოლოგიურ გადაწყვეტილებებში.

განსაკუთრებული ყურადღება უნდა დაეთმოს მცირე და საშუალო საწარმოებისათვის ისეთი ციფრული პროდუქტების შექმნას, რომლებიც ქართულ ენაზე ხელოვნური ინტელექტის გამოყენებას გაამარტივებს და ბიზნესპროცესების ავტომატიზაციას შეუწყობს ხელს.

- ქართული ენის ტექნოლოგიური განვითარება მნიშვნელოვნად არის დამოკიდებული საერთაშორისო სამეცნიერო თანამშრომლობაზე. აუცილებელია აქტიური მონაწილეობა მრავალეროვან კვლევით კონსორციუმებში, საერთაშორისო საგრანტო პროგრამებსა და ღია სამეცნიერო ინიციატივებში.

ასეთი თანამშრომლობა უზრუნველყოფს არა მხოლოდ ფინანსურ და ტექნოლოგიურ მხარდაჭერას, არამედ ცოდნის გაცვლას, საუკეთესო პრაქტიკის გაზიარებასა და ქართული ენის ინტეგრაციას მრავალენოვან ხელოვნური ინტელექტის სისტემებში.

- მომდევნო ათწლეულის განმავლობაში გენერაციული ხელოვნური ინტელექტი, მრავალმოდალური სისტემები და აგენტური (Agentic AI) ტექნოლოგიები კიდევ უფრო მნიშვნელოვან როლს შეასრულებს საზოგადოებრივ ცხოვრებაში. ამ ცვლილებების ფონზე აუცილებელია, რომ ქართული ენა არა მხოლოდ ადაპტირდეს ახალ ტექნოლოგიურ გარემოს, არამედ აქტიურად იყოს წარმოდგენილი მის განვითარებაში. ამ მიზნის მისაღწევად საჭიროა გრძელვადიანი ინვესტიციები ენობრივ რესურსებში, კვალიფიციური ადამიანური კაპიტალის განვითარება, თანამედროვე კვლევითი ინფრასტრუქტურის შექმნა და საერთაშორისო თანამშრომლობის გაძლიერება. მხოლოდ ამ პირობებში იქნება შესაძლებელი ქართული ენის კონკურენტუნარიანობის შენარჩუნება და მისი ეფექტიანი გამოყენება ხელოვნური ინტელექტის ეპოქაში.

დასკვნა

თანამედროვე ციფრული ტრანსფორმაციის პირობებში ენა აღარ წარმოადგენს მხოლოდ კულტურული იდენტობისა და კომუნიკაციის საშუალებას; იგი იქცა ტექნოლოგიური განვითარების, ცოდნის წარმოებისა და ინფორმაციული სუვერენიტეტის ერთ-ერთ განმსაზღვრელ ფაქტორად. ხელოვნური ინტელექტის, ბუნებრივი ენის დამუშავების (Natural Language Processing – NLP), დიდი ენობრივი

მოდელებისა (Large Language Models – LLMs) და მრავალმოდულური სისტემების სწრაფმა განვითარებამ არსებითად შეცვალა ადამიანის, ინფორმაციისა და ტექნოლოგიების ურთიერთქმედების ფორმები. ამ პროცესების ფონზე განსაკუთრებულ მნიშვნელობას იძენს მცირე და საშუალოდ გავრცელებული ენების, მათ შორის ქართული ენის, სრულფასოვანი ინტეგრაცია თანამედროვე ციფრულ ეკოსისტემაში.

წინამდებარე სტატიამ აჩვენა, რომ ქართული ენის ციფრული ტრანსფორმაცია კომპლექსური, მრავალგანზომილებიანი და ინტერდისციპლინური პროცესია, რომელიც მოიცავს ლინგვისტიკას, კომპიუტერულ მეცნიერებას, ხელოვნურ ინტელექტს, მონაცემთა მეცნიერებას, ინფორმაციულ ტექნოლოგიებს, განათლებასა და საჯარო მმართველობას. აღნიშნული მიმართულებების ეფექტიანი ინტეგრაცია წარმოადგენს ქართული ენის ტექნოლოგიური განვითარების აუცილებელ წინაპირობას.

კვლევის თეორიულმა ანალიზმა ცხადყო, რომ თანამედროვე ენობრივი ტექნოლოგიების საფუძველს ქმნის ხარისხიანი ენობრივი რესურსები. ტექსტური კორპუსები, ელექტრონული ლექსიკონები, მორფოლოგიური და სინტაქსური ანოტაციები, პარალელური კორპუსები, მეტყველების მონაცემთა ბაზები და ტერმინოლოგიური რესურსები განსაზღვრავს ხელოვნური ინტელექტის სისტემების სიზუსტეს, ეფექტიანობასა და ფუნქციურ შესაძლებლობებს. შესაბამისად, ქართული ენისთვის ასეთი რესურსების შექმნა და მუდმივი განახლება არა მხოლოდ სამეცნიერო, არამედ სახელმწიფოებრივი მნიშვნელობის ამოცანაა.

სტატიაში განხილულმა საერთაშორისო გამოცდილებამ დაადასტურა, რომ მცირე ენების წარმატებული ციფრული განვითარება დამოკიდებულია გრძელვადიან სახელმწიფო პოლიტიკაზე, ღია მონაცემთა ხელმისაწვდომობაზე, აკადემიური და კერძო სექტორების თანამშრომლობაზე, კვლევითი ინფრასტრუქტურის განვითარებასა და საერთაშორისო პარტნიორობაზე. ესტონეთის, ფინეთის, ისლანდიისა და სხვა ქვეყნების მაგალითები ცხადყოფს, რომ მოსახლეობის სიმცირე არ წარმოადგენს ტექნოლოგიური განვითარების შემაფერხებელ ფაქტორს, თუ არსებობს თანმიმდევრული სტრატეგია და შესაბამისი ინვესტიციები.

ქართული ენის სპეციფიკის ანალიზმა აჩვენა, რომ მისი მორფოლოგიური სირთულე, მდიდარი ზმნური სისტემა, თავისუფალი სიტყვათა რიგი და უნიკალური დამწერლობა ქმნის როგორც გამოწვევებს, ისე შესაძლებლობებს ხელოვნური ინტელექტის სისტემების განვითარებისთვის. ეს თავისებურებები მოითხოვს ისეთი ალგორითმებისა და მონაცემთა რესურსების შექმნას, რომლებიც მორგებული იქნება ქართული ენის სტრუქტურულ და ფუნქციურ მახასიათებლებზე.

ნაშრომში განსაკუთრებული ყურადღება დაეთმო გენერაციული ხელოვნური ინტელექტის განვითარებას და მის გავლენას ქართულ ენაზე. გამოიკვეთა, რომ დიდი ენობრივი მოდელების ხარისხი პირდაპირ არის დამოკიდებული ქართულენოვანი მონაცემების მოცულობაზე, მრავალფეროვნებასა და ხარისხზე. ამიტომ აუცილებელია ეროვნული ტექსტური და ხმოვანი კორპუსების გაფართოება, დიალექტური და ისტორიული მასალების ციფრულიზაცია, სამეცნიერო ტერმინოლოგიის

სტანდარტიზაცია და ღია ლიცენზიით ხელმისაწვდომი მონაცემთა ბაზების განვითარება.

კვლევის საფუძველზე შეიძლება ჩამოყალიბდეს რამდენიმე ძირითადი დასკვნა.

პირველი, ქართული ენის ციფრული ტრანსფორმაცია უნდა განიხილებოდეს როგორც ეროვნული სტრატეგიული პრიორიტეტი, რომელიც უკავშირდება ქვეყნის ტექნოლოგიურ განვითარებას, საგანმანათლებლო სისტემის მოდერნიზაციას, კულტურული მემკვიდრეობის დაცვასა და საერთაშორისო კონკურენტუნარიანობის ზრდას.

მეორე, თანამედროვე ხელოვნური ინტელექტის ეფექტიანი გამოყენება ქართულ ენაზე შეუძლებელია მაღალხარისხიანი ენობრივი რესურსების, სტანდარტიზებული მონაცემებისა და მრავალფეროვანი კორპუსების გარეშე. შესაბამისად, რესურსების შექმნა და მათი ხელმისაწვდომობა უნდა გახდეს ეროვნული ენობრივი პოლიტიკის მნიშვნელოვანი მიმართულება.

მესამე, ქართული ენის ტექნოლოგიური განვითარება მოითხოვს ინტერდისციპლინურ თანამშრომლობას. ლინგვისტების, პროგრამისტების, მონაცემთა მეცნიერების სპეციალისტების, განათლების მკვლევრებისა და საჯარო სექტორის ერთობლივი მუშაობა წარმოადგენს მდგრადი პროგრესის აუცილებელ პირობას.

მეოთხე, უნივერსიტეტების, კვლევითი ცენტრებისა და ტექნოლოგიური კომპანიების თანამშრომლობა მნიშვნელოვნად დააჩქარებს ინოვაციური ქართულენოვანი პროგრამული პროდუქტების, მეტყველების ტექნოლოგიებისა და გენერაციული ხელოვნური ინტელექტის გადაწყვეტილებების შექმნას.

მეხუთე, აუცილებელია ქართული ენის ინტეგრაცია საერთაშორისო კვლევით პროექტებში, მრავალენოვან ენობრივ რესურსებსა და ღია სამეცნიერო ინიციატივებში, რაც გაზრდის ქართული ენის ხილვადობას გლობალურ ციფრულ სივრცეში.

წინამდებარე ნაშრომს აქვს როგორც თეორიული, ისე პრაქტიკული მნიშვნელობა. თეორიული თვალსაზრისით, იგი აერთიანებს ციფრული ტრანსფორმაციის, ხელოვნური ინტელექტისა და ენობრივი ტექნოლოგიების შესახებ არსებულ სამეცნიერო ცოდნას და განიხილავს მას ქართული ენის განვითარების კონტექსტში. პრაქტიკული თვალსაზრისით კი წარმოდგენილია რეკომენდაციები სახელმწიფო პოლიტიკის, უმაღლესი განათლების, სამეცნიერო კვლევისა და ტექნოლოგიური სექტორისათვის, რაც შეიძლება გამოყენებულ იქნეს ქართული ენის ციფრული განვითარების სტრატეგიული დაგეგმვის პროცესში.

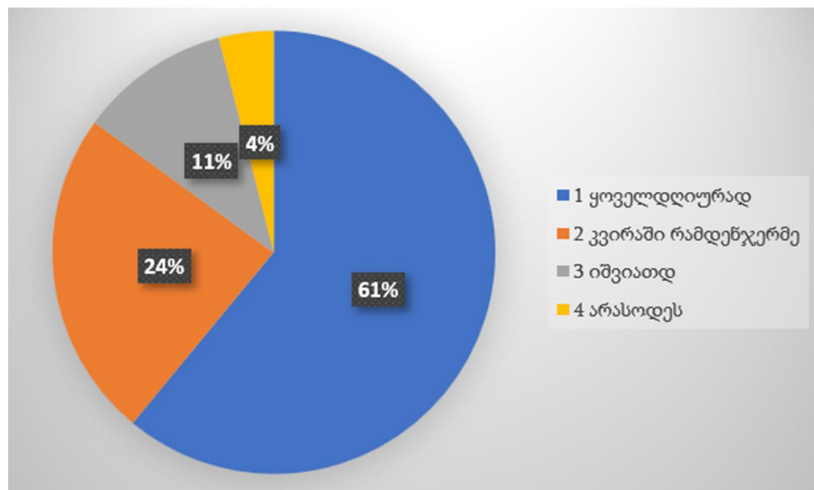
ამავდროულად, კვლევას აქვს გარკვეული შეზღუდვები. ნაშრომი ძირითადად ეყრდნობა თეორიულ ანალიზსა და საერთაშორისო გამოცდილების მიმოხილვას. შესაბამისად, მომავალი კვლევები მიზანშეწონილია დაეფუძნოს ფართომასშტაბიან ემპირიულ მონაცემებს, ქართული ენის კორპუსების რაოდენობრივ ანალიზს, თანამედროვე გენერაციული ხელოვნური ინტელექტის სისტემების შედარებით შეფასებას და მომხმარებელთა გამოცდილების კვლევას. ასევე მნიშვნელოვანია

ქართული ენის სპეციალიზებული დიდი ენობრივი მოდელების შექმნის შესაძლებლობების, ეფექტიანობისა და გამოყენების სფეროების შემდგომი შესწავლა.

საბოლოოდ შეიძლება ითქვას, რომ ქართული ენის ციფრული ტრანსფორმაცია წარმოადგენს არა მხოლოდ ტექნოლოგიურ გამოწვევას, არამედ ქვეყნის სამეცნიერო, კულტურული და სტრატეგიული განვითარების ერთ-ერთ მნიშვნელოვან მიმართულებას. იმ პირობებში, როდესაც ხელოვნური ინტელექტი სულ უფრო მეტად განსაზღვრავს ცოდნის შექმნისა და გავრცელების პროცესებს, ქართული ენის სრულფასოვანი წარმოდგენა ციფრულ გარემოში აუცილებელი წინაპირობაა მისი სიცოცხლისუნარიანობის, საერთაშორისო კონკურენტუნარიანობისა და მომავალი თაობებისთვის მდგრადი განვითარების უზრუნველსაყოფად.

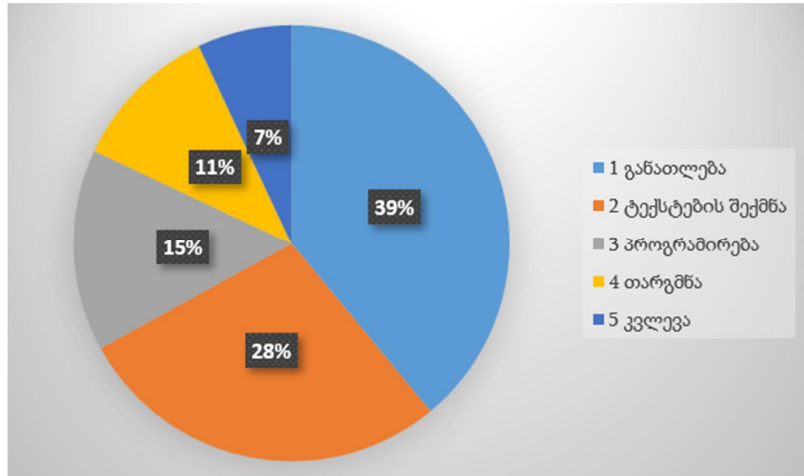
დანართი 1.

1. იყენებთ თუ არა ხელოვნური ინტელექტის სისტემებს ყოველდღიურ საქმიანობაში?



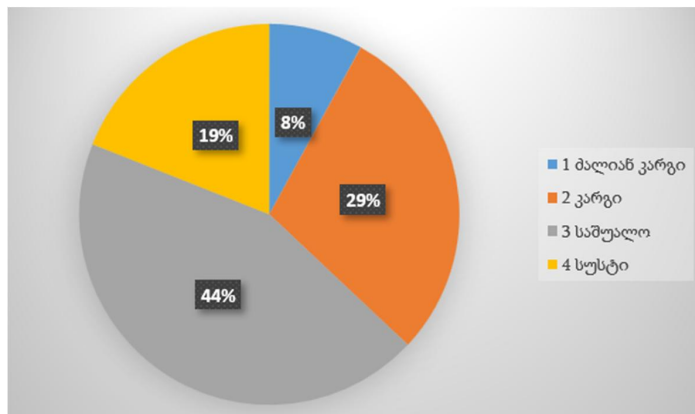
შედეგები მიუთითებს, რომ რესპონდენტთა **85%** რეგულარულად იყენებს ხელოვნური ინტელექტის სხვადასხვა ინსტრუმენტს, რაც ადასტურებს აღნიშნული ტექნოლოგიების ინტენსიურ გავრცელებას საქართველოში.

2. რომელ სფეროში იყენებთ ყველაზე ხშირად AI-ს?



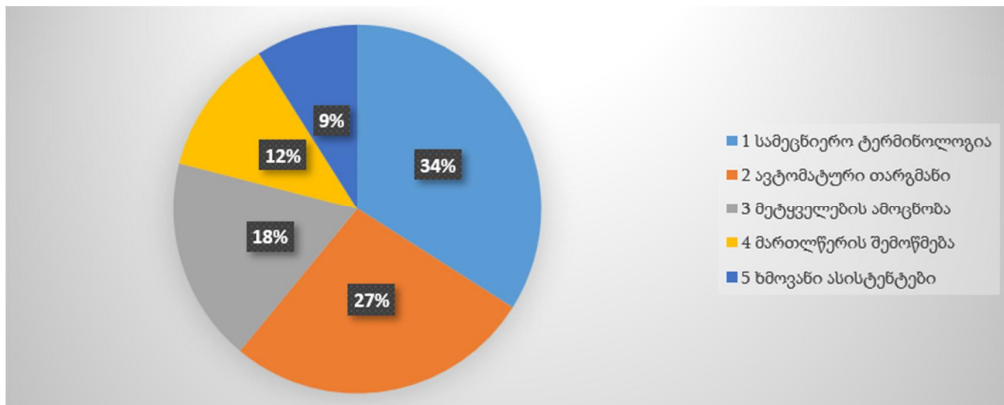
მიღებული შედეგები აჩვენებს, რომ ხელოვნური ინტელექტის ძირითადი გამოყენება დაკავშირებულია საგანმანათლებლო და აკადემიურ საქმიანობასთან.

3. როგორ აფასებთ ქართული ენის მხარდაჭერას თანამედროვე AI სისტემებში?



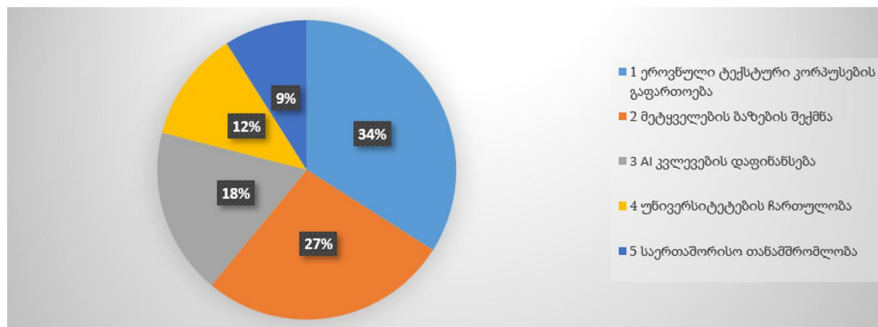
რესპონდენტთა უმრავლესობა მიიჩნევს, რომ ქართული ენის მხარდაჭერა გაუმჯობესებულია, თუმცა ჯერ კიდევ მნიშვნელოვნად ჩამორჩება ინგლისურ და სხვა ფართოდ გავრცელებულ ენებს.

4. რომელი მიმართულებაა ყველაზე პრობლემური?



ყველაზე მნიშვნელოვან გამოწვევად რესპონდენტები ასახელებენ სამეცნიერო ტერმინოლოგიის არასაკმარის მხარდაჭერასა და ავტომატური თარგმნის ხარისხს.

5. რა არის ქართული ენის განვითარების მთავარი პრიორიტეტი?



შედეგები ადასტურებს, რომ რესპონდენტები ყველაზე მნიშვნელოვან მიმართულებად მიიჩნევენ ენობრივი რესურსების განვითარებას.

გამოყენებული ლიტერატურა

- „ქართული ენის ეროვნული კორპუსი“, ქვევ, <https://surl.li/cjupvp>
- მეცნიერებათა ეროვნული აკადემია „მაცნე“ 2024, <https://doi.org/10.48550/arXiv.2405.00710>
<https://arxiv.org/pdf/2405.00710>
- სამოქალაქო ინტეგრაციის და ეროვნებათშორისი ურთიერთობების ცენტრი, „ენობრივი პოლიტიკა საქართველოში: სიტუაციის ანალიზი და კვლევის შედეგები“, 2023, <https://surl.li/ediraz>
- ტაბიძე, მ. „ენობრივი პოლიტიკა“, ეს გვერდი ბოლოს განახლდა 23:18, 30 იანვარი 2024. <https://surl.li/wgmsco>
- ქართული ენის კორპუსი, ილიას სახელმწიფო უნივერსიტეტის ენათმეცნიერების ინსტიტუტი, <https://surl.li/furlsn>

- ქეთევან მჭედლიშვილი, „შედარებითი კორპუსები: შედგენის მეთოდოლოგია დაგამოყენების სფეროები“, კადმოსი. ჰუმანიტარულ კვლევათა ჟურნალი, 2024: 237-253, <https://doi.org/10.32859/kadmos/16/237-253>
- Beso Mikaberidze, Teimuraz Saghinadze, Guram Mikaberidze, Raphael Kalandadze, Konstantine Pkhakadze, Josef van Genabith, Simon Ostermann, Lonneke van der Plas, Philipp Müller, „A Comparison of Different Tokenization Methods for the Georgian Language“, Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024), <https://surl.li/tgrvky>
- Beso Mikaberidze, Teimuraz Saghinadze, Guram Mikaberidze, Raphael Kalandadze, Konstantine Pkhakadze, Josef van Genabith, Simon Ostermann, Lonneke van der Plas, Philipp Müller, „A Comparison of Different Tokenization Methods for the Georgian Language“, Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024) Computer Science > Computation and Language, <https://arxiv.org/abs/2303.08774>
- Computer Science > Computation and Language, OpenAI. (2023). <https://arxiv.org/abs/2303.08774>
- Daniel Gallagher, Gerhard Heyer, Computer Science-Computation and Language, „Targeted Syntactic Evaluation of Language Models on Georgian Case Alignment“, <https://doi.org/10.48550/arXiv.2602.10661>
- David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba O. Alabi, Shamsuddeen H. Muhammad, Peter Nabende, „MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition,“ *EMNLP 2022*. <https://surl.li/pdiwjx>
- David Ifeoluwa Adelani, Md Mahfuz Ibn Alam, Antonios Anastasopoulos, Akshita Bhagia, Marta R. Costa-jussà, Jesse Dodge, Fahim Faisal, Christian Federmann, Natalia Fedorova, Francisco Guzmán, Sergey Koshelev, Jean Maillard, Vukosi Marivate, Jonathan Mbuya, Alexandre Mourachko, Safiyyah Saleem, Holger Schwenk, Guillaume Wenzek, „Findings of the WMT'22 Shared Task on Large-Scale Machine Translation Evaluation for African Languages“, Volume: Proceedings of the Seventh Conference on Machine Translation (WMT), December, 2022. <https://surl.li/yxlqge>
- Davit Melikidze, Alexander Gamkrelidze, „Homonym Sense Disambiguation in the Georgian Language“,
- Emily M. Bender, Batya Friedman, „Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science“, Transactions of the Association for Computational Linguistics, vol. 6, pp. 587–604, 2018. Action Editor: Yuji Matsumoto. Submission batch: 5/2018; Revision batch: 8/2018; Published 12/2018. <https://aclanthology.org/Q18-1041.pdf>
- European Commission „Language Technologies and Multilingualism“, <https://surl.li/wkqvks>, Last update, 23 June 2026

- Fajri Koto, Afshin Rahimi, Jey Han Lau, Timothy Baldwin, „IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP, November 2020, Internationalization (i18n), Making the World Wide Web worldwide!, <https://www.w3.org/International/>
- Irina Lobzhanidze, Erekle Magradze, Svetlana Berikashvili, Anzor Gozalishvili, Tamar Jalaghonia, „Building a Universal Dependencies Treebank for Georgian“, Publisher: Association for Computational Linguistics, Volume: Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024), <https://aclanthology.org/2024.tlt-1.5/>
- Jerome Aondongu Achir, Matthew T Ogedengbe, Joseph Sarwuan, „Preservation of Low Resource Languages through Natural Language Processing: Challenges, Opportunities, and the Case of Tiv“, Jurnal of Basics and Applied Sciences Researcher (JOBASR), Volume 4(2), 2026 DOI: <https://dx.doi.org/10.4314/jobasr.v4i2.21>
- Jurafsky, D. and Martin, J. H., „Speech and Language Processing (3rd ed., draft).“ Stanford University. (2023), <https://surli.cc/gimopn>
- Kamarauli, M. (2024). Enhancement possibilities for the Georgian National Corpus . Caucasus Journal of Social Sciences, 17(1), 142–165. <https://doi.org/10.62343/cjss.2024.245>
- Manning, S. D. and Schutz, H „Fundamentals of Statistical Natural Language Processing“, (1999). <https://surl.li/vuvkns>
- Mariam Kamarauli, „Enhancement possibilities for the Georgian National Corpus“, 2024, <https://surl.li/qywojh> <https://doi.org/10.62343/cjss.2024.245>,
- Nicolas Stefanovitch, Jakub Piskorski, Sopho Kharazi, „Resources and Experiments on Sentiment Classification for Georgian“, Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC 2022), <https://doi.org/10.63317/4ky6so2mrss6>
- OECD, 2024, „Policies, data and analysis for trustworthy artificial intelligence“, <https://oecd.ai/en/>
- Paul Meurer „A Computational Grammar for Georgian“, Conference: Logic, Language, and Computation, 7th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2007, Tbilisi, Georgia, October 1-5, 2007. Revised Selected Papers, https://doi.org/10.1007/978-3-642-00665-4_1
- Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, „The State and Fate of Linguistic Diversity and Inclusion in the NLP World“, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6282–6293 July 5- 10, 2020. <https://surl.li/mhhkhku>
- Proceedings of the 13th Conference on Language Resources and Assessment, European Language Resources Association LREC (2022–2024). <https://lrec-conf.org>
- Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, „On the Opportunities and Risks of Foundation Models,, August 2021, DOI: 10.48550/arXiv.2108.07258 <https://surl.li/lmagxm>
- Rodrigo Agerri, Iñaki San Vicente, Jon Ander Campos, Ander Barrena, Xabier Saralegi, Aitor Soroa, Eneko Agirre, „Give your Text Representation Models some Love: the Case for Basque“,

Publisher: European Language Resources Association, Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020. <https://aclanthology.org/2020.lrec-1.588/>

Steven Bird, Ewan Klein, Edward Loper, „Natural Language Processing with Python“, Publisher: O'Reilly, ISBN: 978-0-596-51649-9 January 2009, <https://surl.li/ysucjv>

Tom. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, „Language Models are Few-Shot Learners“, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, <https://surl.li/ecijgw>

UNESCO, 2021, „Recommendation on the Ethics of Artificial Intelligence“ <https://surl.li/givvwl>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. „Attention Is All You Need“ (2017). <https://arxiv.org/abs/1706.03762>

Digital Transformation of the Georgian Language: The Role of Artificial Intelligence and Language Technologies

Irakli Bochorishvili¹; Tinatin Kiladze²; Ketevan Kveselava³; Edisher Chkareuli⁴

¹Associate Professor of Technical University of Georgia, Academic Doctor of Informatics Engineering; ²Professor of Technical University of Georgian, Doctor of International Relations; ³Professor of Technical University of Georgia, Chief Scientist in the Scientific Research Center of LEPL David Aghmashenebeli National Defence Academy of Georgia, Academic Doctor of Informatics Engineering; ⁴Associate Professor of Technical University of Georgia, Academic Doctor of Mass Communication

Abstract

The rapid development of digital technologies has significantly changed the communicative, educational and information environment of modern society. The integration of artificial intelligence and language technologies has acquired particular importance, creating new opportunities for both global and less common languages. The Georgian language, as a language with a unique script and centuries-old cultural heritage, faces significant challenges in the modern digital era. On the one hand, its full integration into modern information technologies is necessary, and on the other hand, the protection of linguistic identity and norms is necessary.

The article discusses modern trends in the digital transformation of the Georgian language, the role of artificial intelligence in the development of language resources, the capabilities of natural language processing technologies, automatic translation systems, speech recognition and synthesis platforms, the impact of automatic spelling and grammar checking corpora, as well as large language models on the digital development of the Georgian language. Special attention is paid to Georgian language corpora, the need to develop lexical resources and digital infrastructure.

The goal of the study is to analyze the main directions of the digital transformation of the Georgian language, assess the existing challenges, and determine strategic steps that will contribute to increasing the competitiveness of the Georgian language in the global digital space.

The findings indicate that the successful digital transformation of the Georgian language largely depends on the development of comprehensive language resources, open data infrastructure, and sustainable collaboration between academic institutions, government agencies, and the technology sector. In addition, the integration of the Georgian language into modern artificial intelligence systems requires a focused effort in language modeling and increased investment in computational linguistics research.

The article states that the digital existence of the Georgian language is not only a technological issue, but also a matter of cultural sustainability and national strategic development in the era of artificial intelligence.

Keywords: Georgian language, artificial intelligence, language technologies, natural language processing, digital transformation, language resources, large language models.