**საქართველოს უნივერსიტეტი**
*მეცნიერებისა და ტექნოლოგიების სკოლა*
**სადოქტორო პროგრამა: კომპიუტერული მეცნიერებები**

*ხელნაწერის უფლებით*

**დავით ბიტმალკიშევ**

# ხედვის კოგნიტური აღწერის მოდელი ქართულ ენაზე

დოქტორის აკადემიური ხარისხის მოსაპოვებლად
წარმოდგენილი ნაშრომის

სადისერტაციო მაცნე

(სპეციალობა: 0613.1.2)

თბილისი
2026

სადისერტაციო ნაშრომი შესრულებულია საქართველოს

უნივერსიტეტის მეცნიერებისა და ტექნოლოგიების სკოლაში.

სამეცნიერო ხელმძღვანელი: პროფესორი, სერგო ცირამუა

საბჭოს თავმჯდომარე: პროფესორი, მერაბ თოფურია

საბჭოს წევრი: პროფესორი, სულხან სულხანიშვილი

ოფიციალური ექსპერტები:

გარე ექსპერტი - ნანა არაბული საქართველოს ტექნიკურ უნივერსიტეტი, ასოცირებული პროფესორი

გარე ექსპერტი - ნათელა არჩვაძე - თბილისის სახელმწიფო უნივერსიტეტი, ასოცირებული პროფესორი

გარე ექსპერტი - ირაკლი ბაშელეიშვილი - ქუთაისის აკაკი წერეთლის სახელმწიფო უნივერსიტეტი , ასოცირებული პროფესორი

დისერტაციის დაცვა შედგება 2026 წლის 27 თებერვალს, 16:00 საათზე;

მისამართი: საქართველოს უნივერსიტეტი, IV კორპუსი, #519 საკონფერენციო სივრცე;

დისერტაციის გაცნობა შეიძლება საქართველოს უნივერსიტეტის

ბიბლიოთეკაში და უნივერსიტეტის ვებ-გვერდზე: www.ug.edu.ge

სადისერტაციო მაცნე დაიგზავნა 2026 წლის 20 იანვარს;

სადისერტაციო საბჭოს მდივანი - ნათია მანჯიკაშვილი

# მიმოხილვა

თანამედროვე მსოფლიოს ერთ-ერთ ყველაზე მნიშვნელოვან გამოწვევას წარმოადგენს ხელოვნური ინტელექტის მოდელების მრავალენოვანი და კულტურულად ადაპტირებული გამოყენება. მიუხედავად იმისა, რომ გლობალურად მრავალი წარმატებული ვიზუალური აღწერის სისტემა არსებობს (მაგალითად, BLIP2, Flamingo, GPT-4V), ისინი ძირითადად ინგლისურენოვანი მონაცემების საფუძველზეა გაწვრთნილი. ამან გამოიწვია სერიოზული გარღვევა ტექნოლოგიაში, თუმცა, მცირე ენობრივი საზოგადოებები, როგორიცაა ქართული, კვლავაც დაუცველია ხელოვნური ინტელექტის მომსახურებების მიმართ.

ამ კონტექსტში განსაკუთრებულ მნიშვნელობას იძენს „Martha“ — კომპიუტერული ხედვისა და ბუნებრივი ენის გენერაციის ჰიბრიდული მოდელი, რომელიც სპეციალურად ქართულ ენაზე ახორციელებს ვიზუალური გარემოს აღწერას. პროექტი აერთიანებს ორ უმნიშვნელოვანეს კომპონენტს: BLIP2-ის ვიზუალურ ტრანსფორმერს, რომელიც სურათებიდან მაღალი დონის სემანტიკურ ტენსორებს აგენერირებს, და ByT5-ის ენობრივ ტრანსფორმერს, რომელიც მონაცემებს ქართულ ენაზე აღწერს. სწორედ ამ სინთეზის მეშვეობით იქმნება სრულიად ახალი შესაძლებლობა: გარემოს ავტომატური აღქმა და აღწერა ქართულ ენაზე.

„Martha“ არ არის მხოლოდ ტექნოლოგიური ინოვაცია — ის არის სოციალურად და კულტურულად მნიშვნელოვანი პროექტი, რომელიც პასუხობს გლობალურ გამოწვევას: როგორ გავხადოთ ხელოვნური ინტელექტი თანასწორი, მრავალენოვანი და ყველასთვის ხელმისაწვდომი. მისი მნიშვნელობა საქართველოსთვის უდიდესია, რადგან ქმნის ახალი თაობის ქართულენოვან ტექნოლოგიას, ხოლო სიახლე მდგომარეობს იმაში, რომ პირველად განხორციელდა მაღალტექნოლოგიური

მოდელების ჰიბრიდიზაცია ქართულ ენაზე ვიზუალური გარემოს აღწერისთვის.

# შესავალი

## თემის აქტუალობა

კვლევის აქტუალობა რამდენიმე ძირითად გამოწვევას უკავშირდება:

გლობალური უთანასწორობა ენებს შორის – მსოფლიო AI ინდუსტრიაში ინგლისურენოვანი რესურსების დომინაცია მცირე ენებს ჩრდილავს. ამდენად, ქართული ენის ტექნოლოგიური ინტეგრაცია პირდაპირ პასუხობს ციფრული სუვერენიტეტისა და ენობრივი მრავალფეროვნების გამოწვევას.

ინკლუზიურობის დეფიციტი – საერთაშორისო მოდელების უმეტესობა არ არის ადაპტირებული ქართულ ენაზე მუშაობისთვის, რაც მნიშვნელოვნად ზღუდავს მხედველობადაქვეითებული და არასახელმწიფო ენების მომხმარებლების უფლებებს.

ტექნოლოგიური უსაფრთხოება – ადგილობრივ ენებზე მომუშავე სისტემების არსებობა ქვეყნისთვის სტრატეგიული მნიშვნელობისაა, რადგან უზრუნველყოფს მონაცემების ლოკალურ დამუშავებასა და უსაფრთხოებას.

ეკონომიკური პოტენციალი – ქართულ ენაზე მომუშავე AI ტექნოლოგიების განვითარება პირდაპირ უწყობს ხელს სტარტაპ ეკოსისტემის გაძლიერებას და საინფორმაციო ტექნოლოგიების გლობალურ ბაზარზე საქართველოს ჩართვას.

# პროექტის ღირებულება

„Martha“-ს ღირებულება განისაზღვრება რამდენიმე თვალსაზრისით:

ენობრივი დამოუკიდებლობა – პროექტი ქმნის ტექნოლოგიას, რომელიც ქართულ ენაზე მომუშავე მომხმარებლებს საშუალებას აძლევს, მიიღონ ზუსტი, ბუნებრივი და კონტექსტუალურად ადეკვატური აღწერები.

ინკლუზიური ტექნოლოგია – ერთ-ერთი ყველაზე მნიშვნელოვანი ასპექტი არის შშმ პირთა, კერძოდ, მხედველობადაქვეითებულთა დახმარება. სისტემა გარემოს აღწერით აუდიო არხზე გადაცემას უზრუნველყოფს, რითაც ზრდის მათ დამოუკიდებელ ცხოვრებასა და საზოგადოებრივ ინტეგრაციას.

საგანმანათლებლო და კულტურული მნიშვნელობა – ქართულ ენაზე სურათების აღწერის შესაძლებლობა მნიშვნელოვნად წაახალისებს როგორც ქართული მონაცემთა ბაზების შექმნას, ისე ადგილობრივ საგანმანათლებლო და სამეცნიერო რესურსებს.

ბიზნეს-ინოვაციური ღირებულება – სისტემა შეიძლება დაინერგოს სხვადასხვა სექტორში: უსაფრთხოებაში (ვიდეო-მონიტორინგი ქართულ ენაზე), მედიცინაში (ვიზუალური დიაგნოსტიკა და დოკუმენტირება), ტურიზმში (ავტომატური გიდის ტექნოლოგია), მედიასა და განათლებაში.

## *სამეცნიერო კვლევის თეორიული სიახლე*

შემოთავაზებული მეთოდოლოგია, რომელიც აერთიანებს ვიზუალურ ტრანსფორმერებსა და ბაიტურ დონეზე მოქმედ ენობრივ მოდელს (ByT5), ქმნის საიმედო საფუძველს ქართულენოვანი კომპიუტერული ხედვის სისტემების განვითარებისათვის. ნაშრომი მნიშვნელოვანია როგორც თეორიული, ისე პრაქტიკული თვალსაზრისით, რადგან იგი ხელს უწყობს თანამედროვე და ინკლუზიური ციფრული ეკოსისტემის ჩამოყალიბებას.

კვლევის მთავარი სამეცნიერო სიახლე მდგომარეობს BLIP2-ის ვიზუალური კომპონენტის და ByT5 ენობრივი მოდელის ჰიბრიდულ ინტეგრაციაში, რომელიც მიმართულია ქართულენოვანი ტექსტის გენერაციის ამოცანის გადაწყვეტისკენ. კერძოდ:

- ქართულ ენაზე ვიზუალური სცენის აღწერისთვის პირველადაა წარმოდგენილი encoder–decoder არქიტექტურაზე დაფუძნებული ერთიანი სისტემა;
- შემუშავებულია ვიზუალური ემბედინგების პროექციის მექანიზმი, რომელიც უზრუნველყოფს მათ სინქრონიზაციას ByT5 მოდელის დამალულ წარმოდგენებთან;
- შექმნილია სპეციალიზებული ქართულენოვანი ვიზუალურ-ტექსტური (Image Captioning) მონაცემთა ბაზა, რომელიც აერთიანებს გამოსახულებებსა და შესაბამის აღწერებს;
- დასაბუთებულია ბაიტურ დონეზე მომუშავე მოდელის უპირატესობა მორფოლოგიურად მდიდარი ენის შემთხვევაში.

კვლევის მეთოდოლოგია ეფუძნება ექსპერიმენტულ და შედარებით-ანალიტიკურ მიდგომებს, რაც სრულად შეესაბამება დასახულ მიზნებს. კვლევის პროცესში განხორციელდა მონაცემთა ბაზის ფორმირება, მონაცემების წინასწარი დამუშავება, მოდელის არქიტექტურული დაგეგმვა, მისი ტრენინგი და მიღებული შედეგების შეფასება. გამოყენებულია როგორც რაოდენობრივი შეფასების ინსტრუმენტები, ისე სხვადასხვა სცენარში ჩატარებული ექსპერიმენტული ტესტირება.

მეთოდოლოგიური ჩარჩო თანმიმდევრულად და ლოგიკურად ასახავს კვლევის ეტაპებს. ლიტერატურის მიმოხილვა მოიცავს თანამედროვე ვიზუალურ-ენობრივი მოდელების განხილვას და ქმნის მყარ თეორიულ საფუძველს წარმოდგენილი არქიტექტურისთვის, თუმცა

სასურველი იქნებოდა უახლესი საერთაშორისო კვლევების უფრო ფართო და სიღრმისეული შედარებითი ანალიზი.

## სამეცნიერო კვლევის პრაქტიკული სიახლე

„Martha“-ს კვლევითი სიახლე რამდენიმე განზომილებით გამოიხატება:

მოდელების ინოვაციური სინთეზი – პროექტი პირველად ახორციელებს BLIP2-ის ვიზუალური ტენსორების ByT5-ის ენობრივ არხთან სინქრონიზაციას სპეციალურად ქართული ენისთვის.

ბაიტ-დონის დამუშავება – ByT5-ის უნიკალური არქიტექტურა (რომელსაც არ აქვს ტოკენიზატორი) საშუალებას იძლევა, ქართულ ენაზე მაღალი ხარისხის ტექსტი გენერირდეს, რაც სხვა მოდელებისთვის პრობლემურია.

ქართული მონაცემთა ბაზის შექმნა – პროექტის ფარგლებში მუშავდება სპეციალური ქართული სურათ-კაპშენის მონაცემთა ბაზა, რაც პირველი შემთხვევაა ამ მასშტაბით.

მრავალმხრივი გამოყენებადობა – ტექნოლოგია არა მხოლოდ სამეცნიერო სფეროს ემსახურება, არამედ სარგებლიანია საზოგადოებრივ, კომერციულ და სახელმწიფო სერვისებში.

სოციალური ინოვაცია – ინკლუზიური ტექნოლოგიის დანერგვა ქართულ ენაზე ხელს უწყობს თანასწორობის ზრდას და ქმნის რეალურ შესაძლებლობებს შშმ პირებისთვის.

## შედეგების მნიშვნელობა

კვლევის შედეგებს ექნება როგორც ეროვნული, ასევე რეგიონული და გლობალური გავლენა:

საქართველოსთვის – პროექტი ხელს შეუწყობს ქართულ ენაზე მომუშავე ხელოვნური ინტელექტის ეკოსისტემის განვითარებას, სტარტაპებისა და უნივერსიტეტების თანამშრომლობას, მონაცემთა ლოკალური ბაზების შექმნას.

რეგიონისთვის – „Martha“ შეიძლება გახდეს მოდელი სამხრეთ კავკასიისა და მცირე ენების საზოგადოებისთვის, სადაც მსგავსი გამოწვევები არსებობს.

მსოფლიოსთვის – კვლევა ამტკიცებს, რომ ხელოვნური ინტელექტის მრავალენოვანი განვითარება შესაძლებელია მცირე ენებისთვისაც, რაც ენების თანასწორობისა და კულტურული მრავალფეროვნების შენარჩუნებას ემსახურება.

# კვლევის მიზანი

პროექტის მთავარი მიზანია ისეთი კომპიუტერული ხედვისა და ენობრივი გენერაციის ჰიბრიდული სისტემის შექმნა, რომელიც ქართულ ენაზე მოახდენს ვიზუალური გარემოს ზუსტ და ბუნებრივ აღწერას. „Martha“-ს მიზანია ქართული ენის ხელოვნური ინტელექტის ეკოსისტემაში ინტეგრაცია, ინკლუზიური ტექნოლოგიების ხელშეწყობა და ქვეყნისთვის სტრატეგიულად მნიშვნელოვანი ენობრივი დამოუკიდებლობის განმტკიცება.

# კვლევის ამოცანები და ეტაპები

## მონაცემთა ბაზის შექმნა და მომზადება

ამოცანა: ქართული სურათ-კაპშენის მონაცემთა ბაზის აგება (სურათებისა და შესაბამისი ტექსტური აღწერების ფორმირება).

მოსალოდნელი შედეგი: მრავალფეროვანი, სტრუქტურირებული და სანდო მონაცემთა ნაკრები, რომელიც საფუძვლად დაედება მოდელის გაწვრთნას.

BLIP2 ვიზუალური მოდულის ადაპტაცია

ამოცანა: BLIP2 მოდელის გამოყენება ვიზუალური ტენსორების გენერირებისთვის და მისი ინტეგრაცია ქართულენოვან არქიტექტურაში.

მოსალოდნელი შედეგი: გარემოს ვიზუალური ობიექტების, კონტექსტისა და სემანტიკური ურთიერთობების მაღალი სიზუსტით ამოცნობა.

ByT5 ენობრივი მოდულის ოპტიმიზაცია

ამოცანა: ByT5 მოდულის ინტეგრაცია ისე, რომ ვიზუალური ტენსორები ქართულ ტექსტად გარდაქმნას.

მოსალოდნელი შედეგი: ბუნებრივი, გრამატიკულად და სტილისტურად სწორი ქართულენოვანი აღწერების გენერაცია.

მოდელების სინქრონიზაცია და ჰიბრიდული არქიტექტურის შექმნა

ამოცანა: ვიზუალური და ენობრივი მოდულების დაკავშირება ერთიან სისტემად (encoder–decoder სტრუქტურა).

მოსალოდნელი შედეგი: ქართული ენისთვის სპეციალურად მორგებული ჰიბრიდული მოდელი, რომელსაც შეუძლია გარემოს აღწერა.

ინკლუზიური ფუნქციონალის შემუშავება

ამოცანა: მოდელის შედეგების აუდიო ფორმატში გარდაქმნა მხედველობადაქვეითებულთათვის.

მოსალოდნელი შედეგი: პრაქტიკული აპლიკაცია, რომელიც რეალურ სოციალურ გამოწვევებს მოაგვარებს.

ტესტირება, შეფასება და ვალიდაცია

ამოცანა: მოდელის მუშაობის შეფასება სხვადასხვა სცენარებში (საგანმანათლებლო, სამედიცინო, ტურიზმი, უსაფრთხოება).

მოსალოდნელი შედეგი: დადასტურებული სიზუსტე, რეალური გამოყენებადობა და პრაქტიკული ღირებულება.

# მოსალოდნელი შედეგები

ქართული სურათ-კაპშენის მონაცემთა ბაზა.

BLIP2 + ByT5-ის ინტეგრაციაზე დაფუძნებული ჰიბრიდული მოდელი.

ქართული ენაზე გარემოს აღწერის უნარი.

ინკლუზიური გადაწყვეტა შშმ პირთათვის.

ქართული ენის ხელოვნური ინტელექტის ეკოსისტემის გაძლიერება და საერთაშორისო ცოდნის ბაზარზე მისი ინტეგრაცია.

# კვლევის მეთოდოლოგია

## ზოგადი მიდგომა

„Martha“-ს კვლევა ეფუძნება კომპიუტერული ხედვისა და ბუნებრივი ენის დამუშავების თანამედროვე ინტეგრირებულ მიდგომას. პროექტის დიზაინი ეფუძნება encoder–decoder არქიტექტურას, სადაც ვიზუალური ნაწილი (BLIP2) უზრუნველყოფს სემანტიკური ტენსორების გენერირებას სურათებიდან, ხოლო ენობრივი ნაწილი (ByT5) ამ ინფორმაციას ქართულ ენაზე გარდაქმნის.

მეთოდოლოგია მრავალეტაპიანია და მოიცავს: მონაცემთა შეგროვებასა და წინამუშავებას, მოდელების ადაპტაციასა და

სინქრონიზაციას, ფუნქციური მოდულის შექმნას, ტესტირებასა და შედეგების ვალიდაციას.

## კვლევის ეტაპები

მონაცემთა შეგროვება და წინამუშავება

ქართული სურათ-კაპშენის მონაცემთა ბაზის შექმნა (საჯარო და ლოკალური მონაცემების ინტეგრაცია).

ტექსტების წინამუშავება, ორთოგრაფიული და სემანტიკური ხარისხის კონტროლი.

სურათების სტანდარტიზაცია (გადამუშავება, ზომის ოპტიმიზაცია).

მოდელების ადაპტაცია

BLIP2 ვიზუალური encoder-ის გამოყენება, რომელიც უკვე გაწვრთნილია მრავალენოვან სურათ-ტექსტურ მონაცემებზე.

ByT5 decoder-ის ოპტიმიზაცია ქართულ ენაზე, ბაიტ-დონის დამუშავების უპირატესობით.

ჰიბრიდული არქიტექტურის სინქრონიზაცია

ვიზუალური და ენობრივი მოდულების დაკავშირება ერთიან encoder–decoder სისტემაში.

ტენსორების სივრცის ოპტიმიზაცია ისე, რომ ByT5-ის გენერირებული ტექსტი იყოს ბუნებრივი და კონტექსტუალურად შესაბამისი.

შეფასება და ვალიდაცია

მომხმარებელთა ტესტირება (human evaluation) ქართულ ენაზე ტექსტების ბუნებრიობისა და სიზუსტის დასადასტურებლად.

## მეთოდოლოგიის შესაბამისობა მიზნებთან

ეს მიდგომა პირდაპირ პასუხობს პროექტის მთავარ მიზანს — შექმნას ჰიბრიდული მოდელი, რომელიც ქართულ ენაზე აღწერს ვიზუალურ გარემოს. მრავალეტაპიანი დიზაინი უზრუნველყოფს როგორც მეცნიერულ სიზუსტეს (საბაზისო მოდელების სინთეზი და მეტრიკებით შეფასება), ისე პრაქტიკულ შედეგს (ინკლუზიური აპლიკაცია რეალური მომხმარებლისთვის).

## შეზღუდვები და უპირატესობები

შეზღუდვები:

ქართული მონაცემების სიმცირე (შედარებით მცირე კორპუსი ინგლისურთან).

მაღალი გამოთვლითი რესურსების საჭიროება (GPU, დიდი მეხსიერება).

ქართულ ტექსტში მრავალი მორფოლოგიური თავისებურება, რაც გენერაციის ხარისხს ართულებს.

უპირატესობები:

ByT5-ის უნიკალური არქიტექტურა, რომელიც ტოკენიზაციის გარეშე მუშაობს და ქართულ ენას ბუნებრივად ამუშავებს.

BLIP2-ის უნივერსალური encoder, რომელიც მდიდარ ვიზუალურ წარმოდგენებს ქმნის.

მრავალსაფეხურიანი ტესტირება უზრუნველყოფს შედეგების სანდოობასა და გამოყენებადობას.

ინკლუზიურობა და სოციალური გავლენა ტექნოლოგიის დანერგვისას.

# მოსალოდნელი რისკები, წინაღობები და მათი შემცირების გზები

„Martha“-ს პროექტის განხორციელებისას მოსალოდნელია რიგი რისკები და პრობლემები, რომელთა გათვალისწინება და მართვა კრიტიკულად მნიშვნელოვანია.

მონაცემთა სიმცირე და ხარისხი – ქართულ ენაზე სურათ-კაპშენის კორპუსი შეზღუდულია.

შემცირების გზა: არსებული საერთაშორისო მონაცემთა ბაზების ადაპტაცია, crowdsourcing-ით ქართული კაპშენების გენერაცია, ლინგვისტების ჩართვა ხარისხის კონტროლში.

მაღალი გამოთვლითი რესურსების საჭიროება – BLIP2 და ByT5 დიდ რესურსებს მოითხოვს, რაც პროცესს აძვირებს.

შემცირების გზა: ღრუბლოვანი სერვისების (Google Cloud, AWS) გამოყენება, მოდელის პარამეტრების ოპტიმიზაცია და LoRA/quantization მეთოდების დანერგვა. (Ha, Shen, Wallis, Allen-Zhu, Li, Wang, Wang,Chen / ჰა, შენ, ვალისი,ალენ-ჟუ,ლი,ლი,ვანგ,ჩენ. 2021)

ენობრივი სპეციფიკა – ქართული ენის მორფოლოგიური სირთულეები შესაძლოა გენერაციის სიზუსტეს შეაფერხოს.

შემცირების გზა: დამატებითი fine-tuning ქართულ მონაცემებზე, ადამიანის შეფასების (human evaluation) ინტეგრაცია. (Hodosh , Young,Hockenmaier/ჯოდოშ,იანგ,ჰოკერმაიერ. 2013)

ინკლუზიური ფუნქციონალის ტესტირების სირთულე – მხედველობადაქვეითებული პირების ჩართვა შესაძლოა ადმინისტრაციულ და ეთიკურ ბარიერებს წააწყდეს.

შემცირების გზა: თანამშრომლობა ადგილობრივ NGO-ებთან და წინასწარი ეთიკური თანხმობის უზრუნველყოფა.

ფინანსური და დროში გადაცილების რისკი – მაღალი ტექნოლოგიური ღირებულება და რესურსების დაგვიანება.

შემცირების გზა: ეტაპობრივი ბიუჯეტირება, ალტერნატიული დაფინანსების წყაროების მოძიება, პარალელური სამუშაოების დაგეგმვა.

# დასკვნა

პროექტი „Martha“ წარმოადგენს პირველ მცდელობას საქართველოში, ვიზუალური გარემოს აღწერის სისტემა ქართულ ენაზე იქმნება თანამედროვე Encoder–Decoder არქიტექტურის საფუძველზე. პროექტის განხორციელების მეთოდოლოგია აერთიანებს:

მონაცემთა ფართო ბაზის შექმნას (300,000 სურათი და ქართული კაპშენი),

BLIP2 encoder-ის გამოყენებას ვიზუალური წარმოდგენების მისაღებად,

ByT5 decoder-ის ინტეგრაციას ტექსტის გენერაციისთვის, რომელიც ქართულ ენაზე მუშაობს,

Projection layer-ის მათემატიკურ სინქრონიზაციას, რათა ვიზუალური და ენობრივი სივრცეები თავსებადი იყოს,

რეგულარიზაციისა და ოპტიმიზაციის თანამედროვე ტექნიკებს, რომლებიც უზრუნველყოფენ სტაბილურ და ზუსტ ტრენინგს.

შედეგად ვიღებთ მოდელს, რომელსაც შეუძლია სურათების ქართულ ენაზე აღწერა, მაღალი სიზუსტით და ბუნებრივი ენობრივი სტილით.

1. მიღებული მოდელი არის blip2 ვიზუალურ-ენობრივი მოდელის გაუმჯობესებული ვერსია , ინოვაციური მიდგომის გზით, ორი მოდელის სინთეზით.
2. Martha მოდელის გადაწვრთვნა შესაძელებელია ნებისმიერ ენაზე, რომლის მხარდაჭერაც აქვს byt5 დიდ ენობრივ მოდელს, შესაბამის ენაზე მონაცემების შექმნის გზით , როგორც ეს მოხდა ქართული ენის შემთხვევაში. ეს ენებია :
    1. Spanish
    2. French
    3. German
    4. Russian
    5. Chinese
    6. Arabic
    7. Portuguese
    8. Armenian
    9. Azerbaijani
    10. Ukrainian
    11. Turkish
    12. Polish

პროექტის წარმატებით განხორციელება შექმნის პრეცედენტს ქართულ AI კვლევებში. მისი მნიშვნელობა მოიცავს:

ენობრივი ინკლუზიურობა – ქართული ენა პირველად იქნება წარმოდგენილი თანამედროვე ვიზუალურ–ენობრივ მოდელებში.

ინკლუზიური ტექნოლოგიები – მხედველობადაქვეითებული ადამიანებისთვის გარემოს აუდიო აღწერის შესაძლებლობა.

განათლება და მეცნიერება – ახალი ინსტრუმენტი, რომელიც გამოიყენება როგორც სასწავლო, ისე კვლევით პროცესებში.

ტურიზმი და კულტურა – საქართველოს კულტურული მემკვიდრეობის პოპულარიზაცია, ვიზუალური მასალების ქართულენოვანი აღწერებით.

მოსალოდნელი სამომავლო კვლევები

პროექტის დასრულების შემდეგ იგეგმება რამდენიმე მიმართულება:

მულტიმოდალური გაფართოება

ვიდეოს აღწერის შესაძლებლობა (Video Captioning).

აუდიოს ინტეგრაცია (Audio–Visual Captioning).

მულტილინგვური მხარდაჭერა

მოდელის გაფართოება რეგიონულ ენებზე (სომხური, აზერბაიჯანული, ოსური).

კოდ-სვიჩინგის მხარდაჭერა ქართულ–ინგლისური ტექსტებისთვის.

რეალურ დროში captioning

დაბალი ლატენტურობის მოდელის ვერსია მობილური აპლიკაციებისთვის.

ინტეგრაცია სმარტ სათვალეებში („ჭკვიანი სათვალე“ მხედველობადაქვეითებულთათვის).

RAG (Retrieval-Augmented Generation) ინტეგრაცია

გენერირებული კაპშენების გამდიდრება გარე ცოდნის ბაზებით (მაგ. ტურისტული ობიექტების ისტორია).
(Tsiramua,Meladze,Davitashvili,Bitmalkishev,Elbakidze/ჩირამუა,მელაძე,დავითაშვილი,ბიტმალკიშევ,ელბაქიძე. 2025)

მოდელის ოპტიმიზაცია

მოდელის შეკუმშვა (quantization, pruning) მობილურ და edge-დევაისებზე გასაშვებად.

ენერგოეფექტური სწავლება მცირე რესურსების მქონე ლაბორატორიებისთვის.

გამოქვეყნებული პუბლიკაციების ნუსხა:

1. David Bitmalkishev, Sergo Tsiramua, Hamlet Meladze, Tinatin Davitashvili. Analyzing Image Patterns and Generating Text: Advances in Multilingual Vision Language Transformers. Workshop CSIT-2025 on "Large Digital Models and Specific Pattern Analyses", the Institute for Informatics and Automation Problems, Yerevan, May 30-31, 2025.

2. Davit Bitmalkishev, Sergo Tsiramua, Hamlet Meladze, Tinatin Davitashvili, Tatia Elbakidze. Question-Answering System Bazed on AI and NLP Models. Proceedings of the XV Scientific Conference of the Union of Mathematicians of Georgia, Batumi, 1-6 September, 2025. https://gmu.gtu.ge/conferences/wp-content/uploads/2025/08/Conference_GMU_2025.pdf

3. Bitmalkishev Davit, Tsiramua Sergo, Meladze Hamlet, Davitashvili Tinatin, Elbakidze Tatia. AI and NLP Models for Q&A in Georgian. Proceedings of the 15th International Conference on Computer Science and Information Technologies CSIT 2025, Erevan, 2025. https://doi.org/10.51408/csit2025_11

4. Bitmalkishev Davit. Design and Training of a Georgian Vision-to-Text Model Using ViT and ByT5. Proceedings of the South

Caucasus Congerence in Artificial Inteligence, SCCAI2025, Tbilisi, 2025.

5. Sergo Tsiramua, Hamlet Meladze, Davit Bitmalkishev. Analyzing Image Patterns and Generating Text: Advances in Multilingual Vision-Language Transformers. Journal “Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications” issue 4, volume 35, 2025.

**The University of Georgia**

*School of Science and Technology*

**PhD Program: Computer Science**



**Davit Bitmalkishev**

# Cognitive Model of Visual Captioning in Georgian

Thesis submitted for the academic degree of Doctor

# Disertation bulletin

(Speciality: 0613.1.2)

Tbilisi

2026

The dissertation was completed in Georgia University School of Science and Technology.

Scientific supervisor: Professor, Sergo Tsiramua

Chairman of the Board: Professor, Merab Topuria

Board Member: Professor, Sulkhan Sulkhanishvili

Official experts: External Expert - Nana Arabuli, Associate Professor, Georgian Technical University

External Expert - Natela Archvadze - Tbilisi State University, Associate Professor

External Expert - Irakli Basheleishvili - Kutaisi Akaki Tsereteli State University, Associate Professor

The dissertation defense will take place on February 27, 2026, at 16:00;

Address: University of Georgia, Building IV, Conference Space #519;

Dissertation can be found at the University of Georgia in the library and on the university website: www.ug.edu.ge

The dissertation messenger was sent on January 20, 2026;

Secretary of the Dissertation Council - Natia Manjikashvili

## Reviews

One of the most significant challenges in the modern world is the multilingual and culturally adapted use of AI models. While there are many successful visual census systems globally (e.g., BLIP2, Flamingo, GPT-4V), they are mostly trained on English-language data. This has led to a serious breakthrough in technology, however, small language communities like Georgian continue to be vulnerable to AI services.

In this context, "Martha" is a hybrid model of computer vision and natural language generation, which specifically describes the visual environment in Georgian. The project combines two most important components: the BLIP2 visual transformer, which generates high-level semantic tensors from images, and the ByT5 language transformer, which describes the data in Georgian. It is through this synthesis that a completely new opportunity is created: automatic perception and description of the environment in the Georgian language.

"Martha" is not just a technological innovation—it is a socially and culturally significant project that responds to a global challenge: how to make AI equal, multilingual, and accessible to all. Its importance for Georgia is great, as it creates a new generation of Georgian-language technology, and the novelty lies in the fact that for the first time, high-

tech models for describing the visual environment in the Georgian language have been hybridized.

# Introduction

## Topical Relevance

The relevance of the research is related to several main challenges:

Global Disparities Between Languages – The dominance of English-language resources in the global AI industry is overshadowing minor languages. Thus, the technological integration of the Georgian language directly responds to the challenge of digital sovereignty and linguistic diversity.

Lack of inclusivity – Most international models are not adapted to work in the Georgian language, which significantly limits the rights of visually impaired and non-state language users.

Technological security – The presence of systems working in local languages is of strategic importance for the country, as it ensures local data processing and security.

Economic potential – The development of AI technologies working in the Georgian language directly contributes to the strengthening of the startup ecosystem and Georgia's inclusion in the global information technology market.

# Project Cost

The value of "Martha" is determined in several ways:

Linguistic independence – The project creates a technology that allows users working in the Georgian language to receive accurate, natural and contextually adequate descriptions.

Inclusive technology – One of the most important aspects is the assistance of people with disabilities, in particular, the visually impaired. The system provides a descriptive audio channel for the environment, thereby increasing their independent living and social integration.

Educational and cultural significance – The ability to describe images in Georgian will significantly encourage the creation of Georgian databases as well as local educational and scientific resources.

Business-Innovative Value – The system can be implemented in various sectors: security (video monitoring in Georgian), medicine (visual diagnostics and documentation), tourism (automatic guide technology), media and education.

## The Theoretical Novelty of the Scientific Research

The proposed methodology, which combines visual transformers and a byte-level language model (ByT5), establishes a reliable foundation for the development of Georgian-language computer vision systems. The work is significant both theoretically and practically, as it contributes to the formation of a modern and inclusive digital ecosystem.

The main scientific novelty of the research lies in the hybrid integration of BLIP2's visual component and the ByT5 language model, aimed at solving the task of Georgian-language text generation. In particular:

- A unified system based on an encoder–decoder architecture is presented for the first time for visual scene description in the Georgian language;
- A projection mechanism for visual embeddings has been developed to ensure their synchronization with the hidden representations of the ByT5 model;
- A specialized Georgian visual-textual (Image Captioning) dataset has been created, combining images with their corresponding descriptions;

- The advantages of a byte-level model for morphologically rich languages have been theoretically substantiated.

The research methodology is based on experimental and comparative-analytical approaches, fully aligned with the stated objectives. The research process included dataset construction, data preprocessing, architectural design of the model, its training, and evaluation of the obtained results. Both quantitative evaluation tools and experimental testing across various scenarios were employed. The methodological framework consistently and logically reflects the stages of the research.

The literature review discusses contemporary vision-language models and establishes a solid theoretical foundation for the proposed architecture; however, a broader and more in-depth comparative analysis of the latest international research would have been desirable.

## Practical Novelty of Scientific Research

The research novelty of "Martha" is expressed in several dimensions:

Innovative Synthesis of Models – The project is the first to synchronize BLIP2 visual tensors with the ByT5 language channel specifically for the Georgian language.

Byte-level processing – ByT5's unique architecture (which does not have a tokenizer) allows for the generation of high-quality text in the Georgian language, which is problematic for other models.

Creation of a Georgian database – within the framework of the project, a special Georgian image-cap database is being developed, which is the first time on this scale.

Versatile Usability – Technology not only serves the scientific field but is also beneficial in public, commercial, and public services.

Social Innovation – The introduction of inclusive technology in the Georgian language contributes to the growth of equality and creates real opportunities for people with disabilities.

## The Importance of Results

The results of the study will have both national, regional and global impact:

For Georgia – The project will contribute to the development of the artificial intelligence ecosystem working in the Georgian language, cooperation between startups and universities, and the creation of local databases.

For the region, "Martha" can serve as a model for the South Caucasus and the small language community, where similar challenges exist.

For the World – The study argues that multilingual AI can be developed even for small languages, which serves to preserve language equality and cultural diversity.

## Purpose of the research

The main goal of the project is to create a hybrid system of computer vision and language generation that will provide an accurate and natural description of the visual environment in Georgian. The goal of "Martha" is to integrate the Georgian language into the AI ecosystem, promote inclusive technologies, and strengthen strategically important language independence for the country.

## Research Objectives and Stages

Database Creation and Preparation

Task: To build a database of Georgian image-capsing (formation of images and corresponding textual descriptions).

Expected Outcome: A diverse, structured, and reliable dataset that will underpin model training.

BLIP2 Visual Module Adaptation

Task: Using the BLIP2 model to generate visual tensors and its integration into Georgian-language architecture.

Expected outcome: Recognition of visual objects, context, and semantic relationships of the environment with high accuracy.

ByT5 Language Module Optimization

Task: Integration of the ByT5 module in such a way as to convert visual tensors into Georgian text.

Expected result: Generation of natural, grammatically and stylistically correct Georgian-language descriptions.

Synchronization of models and creation of hybrid architectures

Task: To connect visual and language modules into a single system (encoder–decoder structure).

Expected outcome: A hybrid model specially tailored for the Georgian language, capable of describing the environment.

Developing Inclusive Functionality

Task: Convert model results into audio format for the visually impaired.

Expected Outcome: A practical application that will solve real social challenges.

Testing, Evaluation and Validation

Task: To evaluate the model's performance in different scenarios (educational, medical, tourism, security).

Expected Outcome: Proven accuracy, real usability, and practical value.

## Expected Results

Georgian Image-Capshen Database.

A hybrid model based on the integration of BLIP2 + ByT5.

Ability to describe the environment in the Georgian language.

An inclusive solution for people with disabilities.

Strengthening the AI ecosystem of the Georgian language and its integration into the international knowledge market.

## Research Methodology

### General Approach

Martha's research is based on a modern integrated approach to computer vision and natural language processing. The design of the project is based on the encoder-decoder architecture, where the visual part (BLIP2) provides the generation of semantic tensors from images, and the language part (ByT5) converts this information into Georgian.

The methodology is multi-stage and includes: data collection and development, adaptation and synchronization of models, creation of a functional module, testing and validation of results.

## Research Stages

Data Collection and Advancement

Creation of a Georgian image-cap database (integration of public and local data).

Advancement of texts, control of spelling and semantic quality.

Standardization of images (recycling, size optimization).

Adaptation of models

Using a BLIP2 visual encoder already trained on multilingual image-text data.

Optimization of the ByT5 decoder in Georgian with the advantage of byte-level processing.

Hybrid architecture synchronization

Connecting visual and language modules into a single encoder-decoder system.

Optimize the tensor space so that the generated text of ByT5 is natural and contextually relevant.

Assessment and validation

Customer testing (human evaluation) to confirm the naturalness and accuracy of texts in the Georgian language.

## Alignment of the methodology with objectives

This approach directly responds to the main goal of the project — to create a hybrid model that describes the visual environment in Georgian. The multi-stage design ensures both scientific accuracy (synthesis of basic models and evaluation by metrics) and practical results (an inclusive application for a real user).

## Limitations and Benefits

Limitations:

Scarcity of Georgian data (relatively small corpus compared to English).

The need for high computing resources (GPU, large memory).

There are many morphological features in the Georgian text, which complicate the quality of generation.

Advantages:

ByT5's unique architecture, which works without tokenization and processes the Georgian language naturally.

A universal encoder of BLIP2 that generates rich visual representations.

Multi-step testing ensures the reliability and usability of the results.

Inclusivity and social impact in technology implementation.

## Expected risks, obstacles and ways to reduce them

A number of risks and problems are expected during the implementation of the "Martha" project, which are critically important to consider and manage.

The scarcity and quality of the data – the image capture in the Georgian language is limited.

Reduction Path: Adaptation of existing international databases, generation of Georgian caps through crowdsourcing, involvement of linguists in quality control.

The need for high computing resources – BLIP2 and ByT5 require large resources, which makes the process more expensive.

Reduction Path: Using cloud services (Google Cloud, AWS), optimizing model settings, and implementing LoRA/quantization methods. (Ha, Shen, Wallis, Allen-Zhu, Li, Wang, Wang, Chen / Ha, Shen, Wallis, Allen-Zhu, Li, Wang, Chen. 2021)

Linguistic specificity – The morphological complexities of the Georgian language may hinder the accuracy of generation.

The way to mitigate it: Additional fine-tuning on Georgian data, the integration of human evaluation. (Hodosh, Young, Hockenmaier, 2013)

The Difficulty of Testing Inclusive Functionality – Engaging visually impaired individuals may face administrative and ethical barriers.

Path to Mitigation: Collaborating with local NGOs and securing prior ethical consent.

Risk of financial and time delays – high technological cost and resource delays.

Reduction Path: Phased budgeting, finding alternative sources of funding, planning parallel work.

# Conclusion

The project "Martha" is the first attempt in Georgia, the visual environment description system is being created in Georgian on the basis of modern Encoder-Decoder architecture. The methodology of the project implementation includes:

Creating an extensive database (300,000 images and Georgian capsules),

using the BLIP2 encoder to obtain visual representations,

Integration of ByT5 decoder for text generation that works in Georgian,

Mathematical synchronization of the projection layer so that visual and linguistic spaces are compatible,

State-of-the-art techniques of regulation and optimization that ensure stable and accurate training.

As a result, we get a model that can describe images in Georgian, with high accuracy and natural linguistic style.

1. The obtained model is an improved version of the BLIP-2 vision–language model, created through an innovative approach by synthesizing two models.
2. The Martha model can be retrained in any language supported by the ByT5 large language model, by generating data in the corresponding language, as was done in the case of the Georgian language. These languages are:

Spanish

French

German

Russian

Chinese

Arabic

Portuguese

Armenian

Azerbaijani

Ukrainian

Turkish

Polish

The successful implementation of the project will set a precedent in Georgian AI research. Its meaning includes:

Linguistic inclusivity – the Georgian language will be represented for the first time in modern visual-linguistic models.

Inclusive technologies – the ability to provide audio descriptions of the environment for visually impaired people.

Education and Science – a new tool that is used in both educational and research processes.

Tourism and Culture – Promotion of Georgia's cultural heritage with Georgian-language descriptions of visual materials.

Expected Future Studies

After the completion of the project, several directions are planned:

Multimodal Expansion

Video captioning capability.

Audio integration (Audio–Visual Captioning).

Multilingual Support

Extension of the model to regional languages (Armenian, Azerbaijani, Ossetian).

Code-switching support for Georgian-English texts.

Real-time captioning

Low latency model version for mobile applications.

Integration into smart glasses ("smart glasses" for the visually impaired).

RAG (Retrieval-Augmented Generation) integration

Enriching the generated capsules with external knowledge bases (e.g., History of Tourist Objects).
(Tsiramua,Meladze,Davitashvili,Bitmalkishev,Elbakidze/Chiramua,Meladze,Davitashvili,Bitmalkishev,Elbakidze. 2025)

Model optimization

Model compression (quantization, pruning) for running on mobile and edge devices.

Energy-efficient training for laboratories with limited resources.

List of publications published within the framework of the work

David Bitmalkishev, Sergo Tsiramua, Hamlet Meladze, Tinatin Davitashvili. Analyzing Image Patterns and Generating Text: Advances in Multilingual Vision Language Transformers. Workshop CSIT-2025 on "Large Digital Models and Specific Pattern Analyses", the Institute for Informatics and Automation Problems, Yerevan, May 30-31, 2025.

Davit Bitmalkishev, Sergo Tsiramua, Hamlet Meladze, Tinatin Davitashvili, Tatia Elbakidze. Question-Answering System Bazed on AI and NLP Models. Proceedings of the XV Scientific Conference of the Union of Mathematicians of Georgia, Batumi, 1-6 September, 2025. https://gmu.gtu.ge/conferences/wp-content/uploads/2025/08/Conference_GMU_2025.pdf

Bitmalkishev Davit, Tsiramua Sergo, Meladze Hamlet, Davitashvili Tinatin, Elbakidze Tatia. AI and NLP Models for Q&A in Georgian.

Proceedings of the 15th International Conference on Computer Science and Information Technologies CSIT 2025, Erevan, 2025. https://doi.org/10.51408/csit2025_11

Bitmalkishev Davit. Design and Training of a Georgian Vision-to-Text Model Using ViT and ByT5. Proceedings of the South Caucasus Congerence in Artificial Inteligence, SCCAI2025, Tbilisi, 2025.

Sergo Tsiramua, Hamlet Meladze, Davit Bitmalkishev. Analyzing Image Patterns and Generating Text: Advances in Multilingual Vision-Language Transformers. Journal “Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications” issue 4, volume 35, 2025.