Didžiųjų kalbos modelių taikymas vaistų atpažinimui lietuviškuose klinikiniuose tekstuose

Gabrielė Skirmantaitė

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Duomenų mokslo ir skaitmeninių technologijų institutas,
Akademijos g. 4, Vilnius, Lietuva
gabriele.skirmantaite@mif.stud.vu.lt

Santrauka. Lietuviškuose elektroninių sveikatos įrašų tekstuose didelė dalis kliniškai reikšmingos informacijos pateikiama nestruktūrizuotu pavidalu, tad automatinis vaistų ar kitų esybių atpažinimas tampa vis aktualesnis. Tyrime nagrinėtos didžiųjų kalbos modelių galimybės atlikti šį uždavinį žymint vaistus tekste specialiomis žymomis. Eksperimentuose naudoti „Gemini 2.5 Flash" ir „Gemini 2.5 Flash-Lite" modeliai bei skirtingos užklausų formuluotės. Gauti rezultatai parodė, kad tokie modeliai gali būti efektyviai taikomi vaistų atpažinimo uždaviniui, o didžiausią įtaką rezultatams turi užklausos formuluotė ir pateikti anotavimo pavyzdžiai.

Raktiniai žodžiai: įvardytų esybių atpažinimas, didieji kalbos modeliai, natūralios kalbos apdorojimas, klinikiniai tekstai, vaistų atpažinimas.

1 Įvadas

Elektroninis sveikatos įrašas (ESĮ) – tai įrašas, kuriame nuosekliai kaupiama informacija apie paciento sveikatą, surinkta apsilankymų skirtingose sveikatos priežiūros įstaigose metu. Duomenys, saugomi ESĮ, gali būti skirstomi į struktūrizuotus bei nestruktūrizuotus. Struktūrizuotiems duomenims priskiriami tokie ESĮ elementai, kaip paciento svoris ir ūgis, o nestruktūrizuotiems duomenims – klinikiniai užrašai, išrašų aprašymai ar grafiniai vaizdai. Nors struktūrizuotus duomenis galima pakankamai lengvai analizuoti naudojant standartinius statistinius ar mašininio mokymosi metodus, nestruktūrizuoti duomenys gali suteikti daug papildomos ir itin vertingos informacijos [1]. Laisvos formos tekstuose gydytojai detaliai aprašo gydymo eigą, klinikinius sprendimus ir paciento būklės pokyčius. Kadangi didelė dalis informacijos ESĮ saugoma nestruktūrizuotu pavidalu ir jos apimtis didėja, šių duomenų

DOI: https://doi.org/10.15388/LMITT.2026.26

analizė rankiniu būdu reikalauja daug resursų, tad šiam tikslui neretai taikomi natūralios kalbos apdorojimo (NKA) metodai.

Analizuojant klinikinius tekstus, NKA metodai dažnai naudojami informacijos išgavimo (angl. *information extraction*) užduotims – iš nestruktūrizuoto teksto automatiškai išgaunant semantiškai reikšmingą informaciją [2]. Viena iš tokių užduočių yra įvardytų esybių atpažinimas (angl. *named entity recognition*). Šio uždavinio tikslas – automatiškai identifikuoti tekste minimus teksto vienetus, atitinkančius realaus pasaulio esybes konkrečioje taikymo srityje. Medicininiame kontekste iš nestruktūrizuotų ESĮ tekstų išskiriamos esybės, nusakančios kliniškai reikšmingas sąvokas, pavyzdžiui, simptomus, diagnostinius tyrimus ar vaistus. Automatiškai iš klinikinių tekstų išgauti vaistai gali būti naudojami sudarant paciento vartojamų vaistų sąrašus, analizuojant skiriamų vaistų ryšius su diagnozėmis ar tiriant vaistų vartojimo tendencijas medicininiuose duomenyse.

Nors NKA metodų taikymas medicinoje, naudojant nestruktūrizuotus ESĮ duomenis, sudaro sąlygas kurti pažangias diagnostines bei analitines priemones, tebesusiduriama su reikšmingais iššūkiais. Nestruktūrizuoti klinikiniai tekstai yra gydytojų laisvos formos įrašai, tad jų kokybė ir nuoseklumas tiesiogiai priklauso nuo individualių dokumentavimo įpročių. Vartojami skirtingi žodžių sinonimai, pasitaikančios gramatinės klaidos apsunkina automatinį klinikinių tekstų apdorojimą ir tikslų informacijos išgavimą. Papildomus sunkumus kelia plati bei unikali medicinos terminija, sutrumpinimai, vartojami žargonai [3]. Taip pat – ribotas kruopščiai anotuotų duomenų rinkinių prieinamumas, nes anotavimas yra imlus laikui, finansiškai brangus ir dažnai reikalauja specifinių klinikinių žinių. Šių išteklių trūkumas riboja sudėtingesnių NKA modelių kūrimą ir taikymą.

Šie iššūkiai dar labiau išryškėja mažiau paplitusių kalbų, tokių kaip lietuvių kalba, kontekste. Tokios kalbos dažnai pasižymi sudėtingomis gramatinėmis struktūromis, didele žodžių formų įvairove ir kitomis kalbinėmis ypatybėmis, kurios apsunkina NKA metodų taikymą [4]. Be to, dauguma anotuotų, viešai prieinamų duomenų rinkinių yra kuriami anglų ar kitoms plačiai naudojamoms kalboms, todėl mažiau paplitusių kalbų atveju tokie ištekliai yra sunkiau prieinami.

Įvardytų esybių atpažinimo uždaviniai gali būti sprendžiami įvairiais metodais – nuo žodynais ir taisyklėmis paremtų sprendimų, mašininio mokymosi metodų iki hibridinių modelių [5]. Šie metodai gali pasiekti aukštą tikslumą, tačiau dažniausiai reikalauja didelių anotuotų duomenų rinkinių.

Pastaruoju metu NKA srityje vis plačiau taikomi didieji kalbos modeliai (angl. *large language models*), galintys atlikti įvairias teksto analizės ir informacijos išgavimo užduotis be specializuoto modelio mokymo. Tokiems modeliams pateikiamas analizuojamas tekstas kartu su instrukcijomis, aprašančiomis norimą užduotį, ir modelis sugeneruoja atitinkamą išvestį. Valdant modelį pasitelkiant užklausas (angl. *prompts*), itin svarbu jas tinkamai suformuluoti, kadangi nuo to priklauso modelio pateikiamų rezultatų kokybė. Taip pat, modeliams pateikiant kelis pavyzdžius (angl. *few-shot learning*), juos galima adaptuoti specifinei užduočiai neturint didelio anotuotų duomenų rinkinio.

Nepaisant sparčiai augančio didžiųjų kalbos modelių taikymo NKA uždaviniuose, jų naudojimas lietuviškų klinikinių tekstų analizei ir medicininių esybių atpažinimui mokslinėje literatūroje nėra plačiai nagrinėtas. Šio darbo tikslas – ištirti didžiųjų kalbos modelių galimybes atpažinti vaistus lietuviškuose klinikiniuose tekstuose. Darbe analizuojama skirtingų užklausų formuluočių bei kelių pavyzdžių pateikimo įtaka esybių atpažinimo rezultatams.

2 Duomenys

Tyrime naudotas nuasmenintų lietuviškų elektroninių sveikatos įrašų duomenų rinkinys, sudarytas bendradarbiaujant su Vilniaus universiteto ligoninės Santaros klinikų retų ligų specialistais. Įrašų tekstuose aprašoma paciento būklė, diagnozės, atlikti tyrimai bei vartojami vaistai. Duomenų rinkinys buvo sudarytas iš 3410 įrašų, tačiau įvertinus anotavimui rankiniu būdu reikalingus resursus, atrinkti ir suanotuoti 80 kandidatinių įrašų. Atrenkant įrašus buvo siekiama įtraukti skirtingo ilgio tekstus bei tekstus, kuriuose tikėtina yra vaistų paminėjimų ir kuriuose nėra. Taip pat buvo užtikrinta, kad nebūtų įtraukti keli to paties paciento įrašai, taip siekiant išvengti pasikartojančių ar labai panašių tekstų. Atrinkti tekstai buvo suanotuoti pagal iš anksto apibrėžtas anotavimo taisykles. Vaistais buvo laikomi tiek bendriniai veikliųjų medžiagų, tiek komerciniai vaistų pavadinimai, o vitaminai, maisto papildai bei bendros vaistų grupės nelaikyti vaistais.

Iš anotuotų įrašų suformuotas galutinis vertinimo rinkinys, sudarytas iš 50 įvairaus ilgio tekstų, iš kurių 35 įrašuose (70 %) buvo bent vienas vaisto paminėjimas (iš viso 155 vaistų pažymėjimų), o 15 įrašų (30 %) vaistų paminėta nebuvo. Toks rinkinio sudarymas leido vertinti ne tik teisingai aptiktus vaistus, bet ir klaidingus aptikimus tekstuose, kuriuose vaistų nėra.

Papildomai buvo sudarytas 5 anotuotų įrašų rinkinys, naudojamas pavyzdžių pateikimui užklausose. Šį rinkinį sudaro įvairaus ilgio tekstai su

skirtingu vaistų paminėjimų skaičiumi, įskaitant įrašą be vaistų ir įrašą su didesniu jų skaičiumi. Kadangi taikant didžiuosius kalbos modelius įvardytų esybių atpažinimo uždavinį galima atlikti be išankstinio modelio apmokymo, klasikinis duomenų skirstymas į mokymo bei testavimo aibes nebuvo taikomas ir visų eksperimentų rezultatai gauti naudojant tą patį vertinimo rinkinį.

3 Metodika

Vaistų atpažinimo uždavinys šiame tyrime formuluotas kaip teksto anotavimo užduotis. Didžiajam kalbos modeliui pateikiamas vienas klinikinis įrašas ir nurodoma pažymėti vaistus pačiame tekste, įterpiant <med> ir </med> žymas aplink atitinkamas teksto atkarpas. Šis žymėjimo būdas leidžia išsaugoti pradinį teksto kontekstą ir tiksliai nustatyti modelio pažymėtų teksto atkarpų ribas. Tai sudaro galimybę modelio rezultatus tiesiogiai lyginti su anotacijomis teksto atkarpų (angl. *span*) lygmeniu, remiantis tiksliomis teksto pozicijomis.

Eksperimentuose naudoti tos pačios modelių šeimos didieji kalbos modeliai „Gemini 2.5 Flash“ ir „Gemini 2.5 Flash-Lite“, pasiekiami per taikomųjų programų sąsają (API). Šie modeliai pasirinkti atsižvelgiant į prieigos kainą. Kadangi modeliai skiriasi dydžiu ir reikalingais skaičiavimo ištekliais, toks pasirinkimas leido įvertinti modelio dydžio įtaką vaistų atpažinimo uždavinio rezultatams. Visi modelių kvietimai atlikti naudojant deterministinį generavimą (temperature = 0), siekiant sumažinti atsitiktinumą ir užtikrinti rezultatų atkuriamumą.

Modeliams buvo pateikiamos skirtingos užklausų formuluotės. Eksperimentuose naudota bazinė, išsami užklausa su taisyklėmis bei dvi išsamios užklausos su anotuotais pavyzdžiais – pirmoji su 3 pavyzdžiais, o antroji su 5. Nors analizuojami tekstai buvo lietuviški, užklausos formuluotos anglų kalba, atsižvelgus į tai, kad didieji kalbos modeliai mokyti naudojant daugiausiai anglų kalbos duomenis. [6] teigiama, jog modelių veikimas skirtingomis kalbomis priklauso nuo kalbos reprezentacijos mokymo duomenyse, o plačiai paplitusių, daug išteklių turinčių kalbų kontekste, ypač anglų, modeliai paprastai pasiekia geresnius rezultatus nei mažiau paplitusių kalbų atveju.

Bazinėje užklausoje, pavaizduotoje 1 pav. pateikiant trumpą užduoties aprašą, modeliui nurodoma pažymėti vaistų pavadinimus tekste.

```
Task: Extract all medication names mentioned in the following Lithuanian
clinical text.

Return the EXACT same text but wrap every medication name with <med></med>
tags. Do not modify, add, or remove any characters other than inserting the
tags.

If no medications are present, return the text unchanged without any tags.

Clinical text:
{text}
```

1 pav. Bazinė užklausa.

Išsamioje užklausoje papildomai nurodomos anotavimo taisyklės (taisyklių pavyzdžiai pateikiami 2 pav. pabrėžiant, kokie terminai laikomi vaistais, o kokie neturėtų būti žymimi kaip vaistų esybės. Išsamiose užklausose su pavyzdžiais modeliams įkeliami anotuoti tekstai, kuriuose vaistų pavadinimai jau pažymėti. Eksperimentuose naudotos užklausos su trimis ir penkiais anotuotais pavyzdžiais, siekiant įvertinti pavyzdžių skaičiaus įtaką modelių rezultatams.

```
Definition:
MEDICATION refers strictly to:
- A specific drug (brand name), OR
- A specific active pharmaceutical substance
Include:
- Brand names
- Active substances
- Inflected forms exactly as written in the text
Combination rule:
- If medications are written as a combination, tag EACH component separately
Exclude:
- Drug classes or groups
- Vitamins
- Supplements or nutraceuticals
- Medical devices
- Procedures or therapies
- Laboratory markers
- Pharmaceutical forms
- Dosage, frequency, administration route
```

2 pav. Užklausa, su papildomomis taisyklėmis.

Modelių grąžinti tekstai buvo apdoroti siekiant išgauti pažymėtas vaistų teksto atkarpas. Pirmiausia tikrinta, ar modelis įterpė <med> žymas ir grąžino nepakitusį tekstą. Jei modelio atsakymas neatitiko reikalaujamo formato

ir tekstas buvo pakeistas, dokumentas buvo laikomas netinkamo formato. Tokiais atvejais pagrindiniame vertinime šie dokumentai buvo vertinami kaip dokumentai be aptiktų esybių.

Modelių rezultatai vertinti trimis lygmenimis: teksto atkarpų, paminėjimų ir vaistų tipų lygmeniu. Teksto atkarpų lygmeniu modelio pažymėtos atkarpos buvo laikomos teisingomis tik tuo atveju, jei jų pradžios ir pabaigos pozicija tiksliai sutapo su vertinimo rinkinio anotacijomis (angl. *exact match*), o nesutapusios ar perteklinės atkarpos laikytos klaidingais aptikimais. Neaptikti anotacijose pažymėti vaistai laikytini praleistais atvejais. Paminėjimų lygmeniu vertinta, ar modelis aptiko visus vaistų paminėjimus tekste, nepriklausomai nuo tikslių pažymėtų atkarpų ribų. Vaistų tipų lygmeniu vertinta, ar modelis nustatė, kokie unikalūs vaistai minimi dokumente, nepriklausomai nuo jų paminėjimų skaičiaus. Paminėjimų ir vaistų tipų lygmenų vertinimui vaistų pavadinimai buvo normalizuojami, suvienodinant skirtingas linksnių formas ar ištaisant gramatines klaidas. Šiam tikslui sudarytas normalizavimo žodynas, kuriame skirtingos žodžių formos susietos su kanonine vaisto forma.

Vertinimui skaičiuotos preciziškumo (angl. *precision*), atkūrimo (angl. *recall*) ir F1 mikro-vidurkio (angl. *micro*-average) metrikos. Modelių rezultatų patikimumui įvertinti apskaičiuoti 95 % pasikliautinieji intervalai taikant savirankos (angl. *bootstrap*) metodą, pakartotinai sudarant duomenų imtis ir kiekvienai imčiai perskaičiuojant vertinimo metrikas. Šis metodas nereikalauja prielaidos, kad nagrinėjamų rodiklių pasiskirstymas yra normalusis, nes pasikliautinieji intervalai buvo nustatyti iš empirinio imčių pasiskirstymo. Skirtingų modelių ir užklausų rezultatų skirtumų statistiniam reikšmingumui įvertinti taikytas porinis savirankos metodas, poromis lyginant F1 metrikos įverčius tarp skirtingų modelių ir užklausų konfigūracijų.

4 Rezultatai

Eksperimentų rezultatai, pateikti 1 lentelėje, parodė, jog modelių veikimas priklauso nuo užklausos formuluotės. Bazinės užklausos atveju modelio „Gemini 2.5 Flash“ teksto atkarpų atitikimo F1 siekė 0,7458, o „Gemini 2.5 Flash-Lite“ – 0,7921. Pateikus išsamią užklausą su aiškiomis taisyklėmis, „Gemini 2.5 Flash“ F1 padidėjo iki 0,9346, o „Flash-Lite“ – iki 0,8119. Atlikus kokybinę klaidų analizę nustatyta, kad modeliai kaip vaistų esybes dažniausiai pažymėdavo vaistų grupes, vitaminus, maisto papildus bei medicininius

prietaisus. Pateikus tikslius vaistų apibrėžimus, taikytus šiame tyrime, modelių pažymėtos esybės labiau atitiko anotuotus duomenis.

Naudojant užklausą su 3 pavyzdžiais taip pat pasiekti aukšti rezultatai: atitinkamai 0,9245 ir 0,9333. Tačiau padidinus pavyzdžių skaičių iki 5, abiejų modelių F1 sumažėjo – iki 0,8669 („Flash“) ir 0,8947 („Flash-Lite“), tad galima teigti, kad didesnis pavyzdžių skaičius ne visuomet lemia geresnius rezultatus. Atlikus kokybinę rezultatų analizę bei patikrinus pavyzdinių įrašų anotavimo kokybę, tiksli šios tendencijos priežastis nebuvo nustatyta. Pagal „Gemini 2.5 Flash“ modelio anotavimo rezultatus, pateiktus 3 pav. matoma, jog pateikus tris pavyzdžius modelis tinkamai sužymėjo visus vaistus pasirinktame įraše, tačiau pateikus penkis pavyzdžius keli vaistų pavadinimai buvo praleisti (raudona spalva žymi vertinimo rinkinyje pažymėtas esybes, kurių modelis nepažymėjo, o žalia – modelio tinkamai aptiktas esybes).

Užklausos su 5 pavyzdžiais rezultatai:

Gydymo korekcija po tyrimų. - Renoprotekcijai ir lėtinei inkstų ligai gydyti skiriama: tab. Dapagliflozinas (Forxiga) 10 mgx1. - Tęsiamas renoprotekcinis gydymas, proteinurijos korekcija: tab. Telmisartani 80 mgx1, tab. Fosinoprili 20 mgx1. - Vartoti ne mažiau 1,5-2L/d. skysčių - AKS monitoravimas, sumažinta druskos dieta - Baltymų sumažinta dieta iki 1g/kg kūno masės - Dislipidemijos gydymas: tęsti tab. Armolipid (Aterolip) po 1 tab./d. + žuvų taukai po 2 caps./d. - Antrinė hiperurikemija: mitybos korekcija: purinų sumažinta dieta + tab. Allopurinoli 150 mg/d. - Hipofosfateminė dieta - Antibiotikoterapija: tab. Amoxicilini 500 mg x 3 (14 d.)

Užklausos su 3 pavyzdžiais rezultatai:

Gydymo korekcija po tyrimų. - Renoprotekcijai ir lėtinei inkstų ligai gydyti skiriama: tab. Dapagliflozinas (Forxiga) 10 mgx1. - Tęsiamas renoprotekcinis gydymas, proteinurijos korekcija: tab. Telmisartani 80 mgx1, tab. Fosinoprili 20 mgx1. - Vartoti ne mažiau 1,5-2L/d. skysčių - AKS monitoravimas, sumažinta druskos dieta - Baltymų sumažinta dieta iki 1g/kg kūno masės - Dislipidemijos gydymas: tęsti tab. Armolipid (Aterolip) po 1 tab./d. + žuvų taukai po 2 caps./d. - Antrinė hiperurikemija: mitybos korekcija: purinų sumažinta dieta + tab. Allopurinoli 150 mg/d. - Hipofosfateminė dieta - Antibiotikoterapija: tab. Amoxicilini 500 mg x 3 (14 d.)

3 pav. Modelio „Gemini 2.5 Flash“ vaistų atpažinimo rezultatų pavyzdys.

Lyginant modelius tarpusavyje, matyti, kad bazinės užklausos atveju šiek tiek didesnį teksto atkarpų F1 pasiekė „Gemini 2.5 Flash-Lite“ modelis (0,7921), o „Gemini 2.5 Flash“ (0,7458). Tačiau naudojant išsamią užklausą su taisyklėmis geresnius rezultatus pasiekė „Gemini 2.5 Flash“ modelis (F1 = 0,9346), o „Flash-Lite“ F1 siekė 0,8119. Pateikiant užklausas su pavyzdžiais modelių rezultatai skyrėsi nedaug – su 3 pavyzdžiais „Flash-Lite“ pasiekė F1 = 0,9333, „Flash“ – 0,9245, o su 5 pavyzdžiais atitinkamai 0,8947 ir 0,8669. Tai rodo, kad mažesnis modelis kai kuriais atvejais gali pasiekti panašius

rezultatus kaip didesnis modelis, ypač naudojant pavyzdžiais papildytas užklausas.

Vaistų paminėjimų lygmens F1 įverčiai sutapo su teksto atkarpų atitikimo F1, kadangi abi metrikos apskaičiuojamos imant kiekvieną vaisto paminėjimą atskirai – pirmuoju atveju pagal poziciją tekste, o antruoju pagal pavadinimą. Vaistų tipų lygmens F1 įverčiai buvo šiek tiek mažesni, nes šiame lygmenyje vertinami unikalūs vaistų tipai, neatsižvelgiant į jų paminėjimų skaičių dokumente. Todėl keli to paties vaisto paminėjimai daro įtaką paminėjimų lygmens rezultatams, tačiau nedaro įtakos vaistų tipų lygmens rezultatams.

1 lentelė. Vaistų atpažinimo skirtingų modelių ir užklausų tipų F1 metrikos.

Modelis	Užklausos tipas	Teksto atkarpų F1(95 % PI)	Paminėjimų F1 (95 % PI)	Vaistų tipų F1 (95 % PI)
Gemini 2.5 Flash	Bazinė	0,7458 (0,6512–0,8171)	0,7458 (0,6512–0,8171)	0,7305 (0,6311–0,8013)
	Išsami	0,9346 (0,8638–0,9783)	0,9346 (0,8638–0,9783)	0,9274 (0,8615–0,9709)
	+ 3 pavyzdžiai	0,9245 (0,8494–0,9749)	0,9245 (0,8494–0,9749)	0,9143 (0,8426–0,9672)
	+ 5 pavyzdžiai	0,8669 (0,7095–0,9695)	0,8669 (0,7095–0,9695)	0,8520 (0,6932–0,9597)
Gemini 2.5 Flash-Lite	Bazinė	0,7921 (0,6692–0,8824)	0,7921 (0,6692–0,8824)	0,7692 (0,6344–0,8648)
	Išsami	0,8119 (0,6800–0,9028)	0,8119 (0,6800–0,9028)	0,7759 (0,6296–0,8797)
	+ 3 pavyzdžiai	0,9333 (0,8837–0,9684)	0,9333 (0,8837–0,9684)	0,9130 (0,8468–0,9592)
	+ 5 pavyzdžiai	0,8947 (0,8256–0,9447)	0,8947 (0,8256–0,9447)	0,8755 (0,8000–0,9315)

Statistinio reikšmingumo vertinimas (žr. 2 lentelę) parodė, kad tiek išsami užklausa (ΔF1 = 0,1901; 95 % PI 0,1071–0,2848), tiek užklausa su 3 pavyzdžiais (ΔF1 = 0,1799; 95 % PI 0,0956–0,2765) pasiekė statistiškai reikšmingai geresnius F1 įverčius, lyginant su bazine užklausa. Skirtumas laikytinas statistiškai reikšmingu, jei 95 % pasikliautinieji intervalai neapima nulio. Užklausa su 5 pavyzdžiais statistiškai reikšmingai besiskirianti nuo bazinės užklausos nebuvo (ΔF1 = 0,1258; 95 % PI -0,0449–0,2716). Lyginant modelius tarpusavyje nustatytas statistiškai reikšmingas skirtumas naudojant iš-

samią užklausą (ΔF1 = 0,1267), tačiau naudojant užklausą su 5 pavyzdžiais statistiškai reikšmingo skirtumo tarp modelių nenustatyta (ΔF1 = -0,0238). Apibendrinant rezultatus galima teigti, kad didžiausią įtaką vaistų atpažinimo rezultatams turėjo užklausos formuluotė, o modelio pasirinkimas turėjo mažesnę įtaką nei užklausos tipas.

2 lentelė. Modelių ir užklausų rezultatų skirtumų statistinis reikšmingumas.

Palyginimas	F1 skirtumas	95 % PI	p reikšmė	Statistiškai reikšminga
Išsami ir bazinė užklausa (Gemini 2.5 Flash)	0,1901	0,1071–0,2848	< 0,001	taip
+ 3 pavyzdžiai ir bazinė užklausa (Gemini 2.5 Flash)	0,1799	0,0956–0,2765	< 0,001	taip
+ 5 pavyzdžiai ir bazinė užklausa (Gemini 2.5 Flash)	0,1258	-0,0449–0,2716	0,124	ne
Gemini 2.5 Flash ir Gemini 2.5 Flash-Lite (išsami užklausa)	0,1267	0,0458–0,2374	< 0,001	taip
Gemini 2.5 Flash ir Gemini 2.5 Flash-Lite (+ 5 pavyzdžiai)	-0,0238	-0,1949–0,0903	0,838	ne

5 Išvados

Atlikti tyrimai parodė, jog didieji kalbos modeliai gali būti taikomi vaistų atpažinimo uždaviniui lietuviškuose klinikiniuose tekstuose. Tyrimo rezultatai leidžia daryti išvadą, kad didžiausią įtaką modelių veikimui turėjo užklausos formuluotė – naudojant bazinę užklausą teksto atkarpų F1 siekė 0,7458 („Gemini 2.5 Flash“) ir 0,7921 („Gemini 2.5 Flash-Lite“), o pateikus išsamią užklausą su aiškiomis anotavimo taisyklėmis F1 padidėjo atitinkamai iki 0,9346 ir 0,8119. Statistinio reikšmingumo analizė parodė, kad tiek išsami užklausa, tiek užklausa su penkiais pavyzdžiais pasiekė statistiškai reikšmingai geresnius rezultatus nei bazinė užklausa. Tai indikuoja, jog aiškiai suformuluotos anotavimo taisyklės ir užduoties aprašymas yra svarbus veiksnys taikant didžiuosius kalbos modelius informacijos išgavimo uždaviniuose.

Taip pat darytina išvada, kad kelių pavyzdžių pateikimas užklausose gali pagerinti modelių rezultatus, tačiau didesnis pavyzdžių skaičius ne visuomet lems tinkamesnes modelio išvestis. Naudojant užklausą su trimis pa-

vyzdžiais pasiekti aukšti F1 įverčiai (0,9245 ir 0,9333), tačiau padidinus pavyzdžių skaičių iki penkių F1 sumažėjo iki 0,8669 ir 0,8947.

Lyginant modelius tarpusavyje nustatyta, kad mažesnis modelis „Gemini 2.5 Flash-Lite“ kai kuriais atvejais pasiekė panašius ar net kiek didesnius F1 įverčius nei didesnis modelis, ypač naudojant užklausas su pavyzdžiais. Statistiškai reikšmingas skirtumas tarp modelių nustatytas tik naudojant išsamią užklausą, o naudojant užklausą su penkiais pavyzdžiais statistiškai reikšmingo skirtumo tarp modelių nenustatyta. Tai rodo, kad užklausos formuluotė gali turėti didesnę įtaką rezultatams nei modelio dydis ar skaičiavimo ištekliai.

Apibendrinant galima teigti, kad didieji kalbos modeliai yra efektyvus vaistų atpažinimo uždavinio sprendimas lietuviškuose klinikiniuose tekstuose net ir turint palyginti nedidelį anotuotų duomenų rinkinį, tačiau modelių veikimas labai priklauso nuo užklausos formuluotės, pateiktų anotavimo taisyklių ir pavyzdžių. Tyrimo rezultatai rodo, kad tinkamai suformuluotos užklausos gali reikšmingai pagerinti įvardytų esybių atpažinimo rezultatus, todėl užklausų kūrimas yra svarbi didžiųjų kalbos modelių taikymo informacijos išgavimo uždaviniuose dalis.

Padėka

Autorė reiškia padėką prof. Gražinai Korvel ir prof. Rimantei Čerkauskienei už pateiktus duomenis bei už suteiktas konsultacijas, kurios buvo svarbios atliekant tyrimą ir rengiant šį straipsnį.

Literatūra

[1] Tayefi, M., Ngo, P., Chomutare, T., Dalianis, H., Salvi, E., Budrionis, A., & Godtliebsen, F. (2021). Challenges and opportunities beyond structured data in analysis of electronic health records. WIREs Computational Statistics.

[2] Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Naveed, A., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., & Liu, H. (2018). Clinical information extraction applications: A literature review. Journal of biomedical informatics, 77, 34–49.

[3] Savova, G. K., Danciu, I., Alamudun, F., Miller, T., Lin, C., Bitterman, D. S., & Tourassi, G. (2019). Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. Cancer research, 79(21), 5463–5470.

[4] Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. Natural Language Processing, 31(2).

[5] Liu , S., Tang, B., Chen, Q., & Wang X. (2015). Drug Name Recognition: Approaches and Resources. Information, 6 (4), 790–810.

[6] Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., Li, M., Che, W., & Yu, S. P. (2025). A survey of multilingual large language models. Patterns, 6(1).