Prasmės problema didžiųjų kalbos modelių veikime: filosofinis tyrimas.

Ar yra pagrindo teigti, jog Wittgensteino prasmės kaip vartojimo teorija susijusi su didžiųjų kalbos modelių veikimo principais?

Turinys

**1. Įvadas:**

* Tema ir jos aktualumas (AI ir filosofija).
* Kodėl pasirinkai Wittgensteiną?
* Tikslas: parodyti, kaip LLM'ai praktikoje atitinka Wittgensteino kalbos vartojimo teoriją.

**2. Wittgensteino teorija trumpai:**

* Reikšmė kaip vartojimas. 🡪 pradinė ir pirminė sąvoka, kuri duoda starta betkokiam mąstymui ar prielaidavimui apie šių dviejų dalykų ryšį. Jį papildo vėlesni dėmenys (kalbos žaidimai, šieminynis panašumas).
* Kalbos žaidimai. 🡪 specializuoti kalbos modeliai, geresnis suvokimas kontekste, kalbos žaidimai kaip semantinės erdvės vietos, kampučiai, besiskiriantys savo apimtimi, gilmumu (pvz. Skiriasi paliepimų kalbos žaidimas nuo kvantinės mechanikos kalbos žaidimo, fizikos ar filosofijos kalbos žaidimų).
* Šeimyninis panašumas.

**3. LLM veikimo principai:**

* Mokymosi duomenys.
* Dėmesio mechanizmas.
* Semantinių ryšių modeliavimas.
* GPT vs Bert? Tiesiog parodymui

**4. Palyginimas:**

* Ar LLM‘ai dalyvauja kalbos žaidimuose?
* Kaip jų veikimas parodo „reikšmę kaip vartojimą”?
* Kur kyla ribos (LLM vs žmogus)?

**5. Kritinė refleksija:**

* Ką tai sako apie kalbos prigimtį?
* Ar LLM’ai padeda suprasti Wittgensteiną geriau ar jį paneigia? O kaip atvirkščiai?

**6. Išvados:**

* Ką galima teigti apie kalbą, reikšmę ir DI?

# Įvadas

Wittgensteinas siūlė atsisakyti idėjos, kad žodžių reikšmės yra fiksuotos, vidinės ar loginės – ir vietoj to kalbėti apie reikšmę kaip vartojimą tam tikroje kalbinėje praktikoje. Tokios sąvokos kaip kalbos žaidimai, gyvenimo forma ir šeiminiai panašumai leidžia matyti, kaip kalba yra veikia, kurios prasmė priklauso nuo konteksto, funkcijos ir socialinės aplinkos.

Techonologijų, ypač dirbtinio intelekto pažanga per pastaruosius metus pravėrė naujas duris ir filosofijoje. Tokių didžiųjų kalbos modelių (*angl.* large language models, LLM) kaip ChatGPT ar BERT veikimo principai ir jų sąryšiai yra migloti, sunkiai įkandami. Šie modeliai, remdamiesi milžiniškais kalbos duomenų kiekiais, geba generuoti tekstą, kuris neretai atrodo prasmingas, tinkamas kontekstui ir funkciškai panašus į žmogaus kalbą. O prie supratimo gryninimo neprisidėda ekonominių interesų turintys individai, ypač iš to besipelnančios įmonės. Wittgensteino manymu, filosofija atsako į klausimus juos panaikindama (CITATA). Todėl norėčiau būtent tam ir pasiryžti. Šio darbo tikslas parodyti ar prasmės kaip vartojimo koncepcija randama šiuolaikinių LLM, tokių kaip GPT ar BERT, veikime, o jei randama, tai koks yra santykis tarp jo ir prasmės kaip vartojimo sampratos, kurią siūlo vėlyvasis Wittgensteinas. Kokiu pagrindu ir kaip galime įmatyti vitgenšteinišką prasmės sampratą LLM‘ų veikime? Šio klausimo pagrįstumas kyla stebint šių technologijų tobulėjimo spartą ir įtikinančius rezultatus praktikoje [ĮTERPTI STRAIPSNIŲ APIE JŲ TOBULĖJIMUS, KĄ GALI DABA DARYT, KAIP PLAČIAI NAUDOJAMI] – kadangi žmonės susišneka tarpusavyje, paliepia vienas kitam, tai į tokia padėtį papuolus LLM‘ams, žodžių prasmės kaip vartojimo supratimas tampa labai įtikinantis siekiant žinojimo. Susišnekėjimas ir, tam tikra prasme, bendradarbiavimas veda prie klausimo apie šių technologijų dalyvavimą kalbos žaidimuose bei tai, kaip pasirodo šeimyninių panašumų (*angl.* family resemblence) koncepcija LLM‘ų „supratimo” galimybėse. Šiame darbe nesiekiama atsakyti, ar šie dirbtinio intelekto modeliai „supranta”, „mąsto” ar „suvokia reikšmę” tokiu būdu, kaip tai daro žmonės. Mano manymu, tokie svarstymai neturint suvokimo struktūros, žodyno ar kalbos žaidimo taisyklių šiai temai yra pernelyg ankstyvi, spekuliaciniai. Vietoj to, tyrimas sutelktas į kalbos vartojimą ir tai, ar šie modeliai gali būti filosofiškai svarbūs, nagrinėjant reikšmės vartojimo kontekstus.

Prieš žengiant prie šiuolaikinių technologijų analizės, būtina išsamiai panagrinėti paties Wittgensteino kalbos teoriją, jos pagrindines sąvokas ir jų tarpusavio ryšius.

# Reikšmė kaip vartojimas.

Read dictionary and use something from there. pradinė ir pirminė sąvoka, kuri duoda starta betkokiam mąstymui ar prielaidavimui apie šių dviejų dalykų ryšį. Jį papildo vėlesni dėmenys (kalbos žaidimai, šieminynis panašumas).

Vienas esminių Wittgensteino vėlyvosios filosofijos teiginių – tai, kad žodžio reikšmė yra jo vartojimas kalboje. „žodžio reikšmė yra jo vartojimas kalboje” (1997, FT §43). Šis požiūris griauna tradicinę idėją, jog žodžių reikšmės yra stabilūs objektai ar loginiai atitikmenys [CITATION FROM RUSSELL, FREGE, AUGUSTUS NEEDED]. Reikšmė, Wittgensteino manymu, nėra tai, kas slypi „už” žodžio ar paslėpta mūsų sąmonėje, greičiau tai – kas atsiskleidžia per praktinį kalbos vartojimą tam tikrame kontekste. Norint suprasti kalbą, reikia stebėti, *kaip* ir *kur* ji veikia, o ne ieškoti jos loginės struktūros ar esminių dėsnių, „negalvok, o žiūrėk!” (FT §66)

Tokią poziciją detaliai komentuoja ir Schroederis, kuris primena, kad net žodynai, kurių tikslas yra apibrėžti žodžių reikšmes, remiasi ne teorinėmis konstrukcijomis, bet praktiniu žodžių vartojimu, ypač tarp išsilavinusių kalbos vartotojų. [CITATION NEEDED] Žodynų sudarytojai fiksuoja ne „tikrąsias” reikšmes, bet tai, *kaip* žodžiai faktiškai vartojami – kokiame kontekste, kokiomis nuotaikomis, kokiems objektams ar žmonėms apibūdinti. [CITATION NEEDED]

Reikšmė nėra atskirta nuo kalbos vartotojo. Schroederio pastebėjimu, vaikai neišmoksta „žodžių”, bet išmoksta, kaip jais naudotis – t. y. perima kalbos žaidimų taisykles [CITATION NEEDED]. Jie geba atskirti, kada galima pasakyti vieną ar kitą žodį, kam jis skirtas, kokiose situacijose jis priimtinas. Toks žinojimas nėra loginė ar semantinė žinių bazė – tai praktinis gebėjimas veikti kalboje, kuris sudaro reikšmės pagrindą. Tai leidžia teigti, kad kalbos supratimas yra dalyvavimas, o ne turėjimas „vidinio žodyno”. Prasmė atsiranda iš kalbinio veiksmo, kurį galima stebėti ir kuriame dalyvaujame nuo vaikystės.

[IS NEEDED?] Wittgensteino požiūris leidžia kitaip permąstyti ir „nesusipratimo” momentus. Pvz., kai bakalauro studentas vartoja sudėtingą akademinį žodyną nesuprasdamas jo tikros paskirties, klausytojas gali tai atpažinti [BŪTENT, ČIA GALIMA SUSIETI SU AI, KOKĮ PAVYZDĮ SU AI HALIUCINACIJOMIS]: studentas nesugeba atskirti, kuo jo vartojamas terminas skiriasi nuo kitų, panašių. Tai reiškia, kad jis neturi *veiksmingo vartojimo suvokimo* – jo žinojimas yra paviršinis, nesuvoktas. Šis pavyzdys parodo, kad tikras žodžio reikšmės supratimas pasireiškia ne per gebėjimą jį pakartoti, bet per gebėjimą jį *tinkamai naudoti*.

[IS NEEDED?] Galų gale, Schroederis ir kiti interpretuotojai pabrėžia, kad Wittgensteinas nekūrė naujos reikšmės teorijos, o veikiau bandė išsklaidyti filosofines painiavas, kylančias iš netinkamo kalbos supratimo. Kaip jis rašo: *„*Filosofija yra kova prieš mūsų proto pakerėjimą mūsų kalbos priemonėmis*”* (FT §109). Vadinasi, reikšmės analizė jam nėra teorinis apibrėžimas, bet praktinis tyrimas, kaip žodžiai veikia mūsų gyvenimo formose, ir kaip mūsų nesusikalbėjimai dažnai kyla ne dėl faktų stokos, o dėl netinkamo žodžių vartojimo supratimo.

[NEEDS MORE AUTHORS, ONE FROM DICTIONARY, OTHER FROM AI ARTICLES]

## Kalbos žaidimai.

specializuoti kalbos modeliai, geresnis suvokimas kontekste, kalbos žaidimai kaip semantinės erdvės vietos, kampučiai, besiskiriantys savo apimtimi, gilmumu (pvz. Skiriasi paliepimų kalbos žaidimas nuo kvantinės mechanikos kalbos žaidimo, fizikos ar filosofijos kalbos žaidimų).

„Atsietai nuo žaidimo *nėra* jokių vardų” (FT 49)

Kalbos žaidimų sąvoka (*angl.* language-games) yra neatskiriama nuo reikšmės kaip vartojimo[CITATION]. Šie du elementai nėra atsitiktinai susiję – kalbos žaidimai yra būtent ta terpė, kurioje vyksta kalbos vartojimas, o kartu ir kuriama reikšmė[CITATION]. Kalbos žaidimų daugiskaita pabrėžia, kad nėra vieno bendro, universalaus kalbos naudojimo būdo – vietoj to egzistuoja įvairios praktikos, kuriose žodžiai įgauna skirtingas prasmes priklausomai nuo veikimo konteksto. [CITATION]

Wittgensteinas rašo: *„Kalbos žaidimas apima kalbos naudojimą ir veiksmus, su kuriais ji susijusi.”* (PI §7) Tai reiškia, kad kalba nėra atskira nuo mūsų kasdienybės: kiekvienas jos panaudojimo atvejis yra tam tikros socialinės veiklos dalis. Vartojimas čia apima ne tik gramatinius sakinius, bet ir jų paskirtį: įsakymas, prašymas, aprašymas, pasakojimas, užuomina – visa tai yra skirtingi kalbos žaidimai .Tokiu būdu reikšmė kyla iš funkcinės, ne loginės sistemos, o žodžių supratimas tampa priklausomas nuo mūsų gebėjimo orientuotis konkrečiose vartojimo situacijose.

Svarbu tai, kad kalbos žaidimai neturi griežtų ribų ar išbaigtų taisyklių – kaip ir šeimyninio panašumo atveju (*family resemblances*, PI §67), žaidimai jungiasi per panašumus, bet ne per vieną bendrą bruožą. Ši kalbos fragmentacija Wittgensteinui leidžia atmesti tiek loginę-semantiškai griežtą kalbos teoriją, tiek bet kokį bandymą kalbą „išgryninti” iki vienos reikšmės sampratos. Anot H.-J. Glock, *„kalbos žaidimų sąvoka atspindi Wittgensteino įsitikinimą, kad kalbos reikšmės priklauso nuo įvairių vartojimo formų, kurios yra ne sumažinamos iki vieno bendro pavadinimo, bet susijusios per įvairius funkcinius panašumus.”* (Glock, 1996, p. 151)

Dėl šios priežasties, kalbos žaidimai yra būdas stebėti reikšmės susidarymą praktikoje, o ne teorinis modelis. Kaip jau parodyta ankstesniame skyriuje apie reikšmę kaip vartojimą, žodynai, semantinės analizės ar parafrazės nėra pakankamos, kad perteiktų žodžio prasmę – ją įmanoma suprasti tik dalyvaujant tame žaidime, kuriame tas žodis naudojamas. Tai reiškia, kad tiek kalbos taisyklių mokymasis, tiek kalbinis supratimas yra įsitraukimas į socialinį veiksmą, kurio taisyklės dažnai nėra aiškiai išsakytos, bet yra perimamos per praktiką.

Wittgensteinas teigia, kad *„kalbos reikšmių neįmanoma paaiškinti abstrakčiai; jas galima parodyti tik per jų naudojimą”* (PI §120). Kalbos žaidimų samprata būtent tai ir daro: ji leidžia atpažinti įvairias kalbos naudojimo formas, padeda demistifikuoti painiavą, kai žodis vartojamas už jam įprasto konteksto ribų, ir parodo, kad nesusipratimai filosofijoje dažnai kyla dėl kalbos žaidimų maišymo – kai vieno žaidimo taisyklės taikomos kitam. “Filosofija yra kova prieš mūsų proto pakerėjimą mūsų kalbos priemonėmis” (FT 109).

Ši sąvoka taip pat leidžia analizuoti kalbą įvairiuose diskursuose – pavyzdžiui, moksliniuose, estetiniuose ar kasdieniuose – kaip savas taisykles turinčius žaidimus, kuriuose žodžių reikšmės „gyvena”. Toks požiūris leidžia ne tik filosofiškai reflektuoti apie kalbą, bet ir, kaip matysime vėlesniuose skyriuose, pritaikyti šią analizę kalbant apie šiuolaikinius dirbtinio intelekto kalbos modelius, kurie patys įsitraukia į tam tikras žaidimų formas, tačiau be žmogiškų gyvenimo formų pagrindo.

Wittgensteino kalbos žaidimų sąvoka yra tiesioginė ir neišvengiama reikšmės kaip vartojimo teorijos išplėtra. Jei reikšmė yra tai, kaip žodis vartojamas kalboje, tai neišvengiamai kyla klausimas: kokiuose kontekstuose ir pagal kokias taisykles tie žodžiai vartojami? Būtent į šį klausimą atsako kalbos žaidimų samprata. Kalbos žaidimas yra tam tikra veikimo forma, kurioje kalbiniai ženklai įgyja prasmę, nes jie yra naudojami pagal tam tikras „žaidimo” taisykles – konvencijas, tikslus, socialinius lūkesčius. Tokie žaidimai gali būti labai įvairūs – nuo kasdienės kalbos („Uždaryk langą!”), iki mokslinių teiginių („Elektronas turi sukamąjį momentą”) ar net meninių išraiškų. Kiekvienas jų reikalauja skirtingo vartojimo būdo, todėl ir žodžių reikšmės skiriasi ne dėl to, kad žodžiai būtų iš esmės kitokie, bet todėl, kad jie įtraukti į skirtingas žaidimų struktūras.

Taigi, kalbos žaidimai sukuria reikšmės galimybės erdvę – tai yra, jie apibrėžia, kada tam tikras žodis ar posakis turi prasmę, o kada jis netenka savo veikimo galios. Pvz., „argumentas” filosofijos kalbos žaidime reiškia loginę struktūrą, o kasdienėje kalboje – nesutarimą. Žodžio „reikšmė” šiuose dviejuose žaidimuose bus labai skirtinga, bet nė viena jų nėra „klaidinga” – jos abi prasmingos savo žaidimo kontekste. Ši įžvalga yra kertinė Wittgensteino kritikai prieš klasikinę reikšmės teoriją, kuri ieškojo vieno, universalaus reikšmės apibrėžimo. Priešingai, Wittgensteinas rodo, kad reikšmė nėra loginė esmė, bet daugialypė funkcija, priklausanti nuo žaidimo taisyklių, kurių dažnai net negalima tiksliai verbalizuoti – jos „gyvenamos”, o ne formalizuojamos.

Šiuo požiūriu kalbos žaidimai yra tarsi scenos, kuriose kalbos vienetai įgyja vaidmenis. Tai leidžia filosofiškai paaiškinti, kodėl tas pats žodis skirtinguose kontekstuose turi kitokį svorį, paskirtį ar net poveikį – nes jis „vaidina” kitame žaidime, laikosi kitų normų. Kalbos žaidimų idėja taip pat padeda suprasti, kodėl nesusikalbėjimai dažnai kyla ne dėl gramatinių ar semantinių klaidų, bet dėl to, kad dalyviai nesutaria, kokį žaidimą jie žaidžia. Tokiu būdu reikšmės ir kalbos žaidimo sąsaja tampa neatskiriama: kalbos žaidimai įrėmina, apriboja ir įgalina reikšmių atsiradimą. Be žaidimo – nėra vartojimo, o be vartojimo – nėra reikšmės.

## Šeimyninis panašumas.

Šeimyninio panašumo (*angl.* family resemblance) sąvoka yra būtinas kalbos žaidimų ir reikšmės kaip vartojimo sampratos papildymas. Jei reikšmė kyla iš vartojimo, o vartojimas vyksta įvairiuose kalbos žaidimuose, tai neišvengiamai kyla klausimas: kas sieja šiuos žaidimus tarpusavyje? Kodėl mes vis dar kalbame apie „kalbą” ar „reikšmę”, jei jos pasireiškia tokiais skirtingais būdais? Šeimyninis panašumas atskleidžia, kad ryšiai tarp kalbos formų yra ne loginiai ar esminiai, bet panašūs į šeimos narių bruožų panašumus.

Wittgensteinas šią idėją pristato per žaidimų pavyzdį: „Pavyzdžiui, panagrinėkime procesus, kuriuos vaidiname „žaidimais“, turiu mintyje šachmatų žaidimą, lošimą kortomis, žaidimą su kamuoliu, sporto žaidimus ir t.t. Kas bendra jiems visiems? – Nesakyk: „Jiems *turi būti* kažkas bendra, kitaip jie nesivadintų „žaidimais“ – bet *įsižiūrėk*, ar jie turi ką nors bendra“ ir toliau paaiškina, jog įsižiūrėję pamatysime „ne tai, kas jiems *visiems* bendra, bet panašumus, giminingumus, ir jų yra labai daug“ (FT §66). Kitaip tariant, daugelis „aiškių“ ir „apibrėžtų“ sąvokų iš tikrųjų neturi vieno bendro esminio bruožo, vietoj to jos yra viena kitą užklojančių panašumų tinkle (ibid.), kaip šeimos nariai gali turėti panašų nosies pavidalą, akių spalvą ar ūgį, bet nė vieno bruožo, kuris būtų bendras visiems.

Svarbu tai, kad šeimyninis panašumas nėra kompromisas ar nepilnas apibrėžimas. Pasak Glocko (1996, 120 p.), šeimyninių panašumų sąvoka yra svarbi pasisakant prieš „požiūrį, kad visiems sąvokos atvejams turi būti kažkas bendro, kas paaiškina, kodėl jie priskiriami šiai sąvokai“ ir „kad vienintelis tinkamas arba teisėtas žodžio paaiškinimas yra analitinis apibrėžimas, kuriame nustatomos būtinos ir pakankamos jo taikymo sąlygos“. Kitaip tariant, dažnai neryškus vaizdas yra kaip tik tai, ko mums reikia (FT §71).

Šeimyninis panašumas taip pat padeda suprasti, kodėl kalbos mokymasis yra galimas be išankstinio visų taisyklių žinojimo. Vaikai neišmoksta apibrėžimų – jie perima vartojimo modelius, kurie formuojasi per šeiminius panašumus su jau žinomais atvejais (plg. FT §73) . Schroeder pabrėžia, kad šeimyninis panašumas parodo, kaip kalbos kompetencija formuojasi per praktikos įgūdžius, o ne per teorinį mokymąsi (2006, 169 p.). Būtent todėl vaikai gali pradėti vartoti žodį „ačiū”, „pienas“ ir kt. žinodami tik kelis pavyzdžius.

Galų gale, šeimyninis panašumas ne tik papildo reikšmės kaip vartojimo ir kalbos žaidimų sampratas, bet ir suteikia joms lankstumą, reikalingą analizuojant sudėtingas kalbos praktikas. Reikšmės kaip vartojimo kontekste šeimyninis panašumas leidžia suprasti, kodėl tas pats žodis gali funkcionuoti skirtinguose žaidimuose išlaikydamas tam tikrą tapatumą. Pavyzdžiui, žodis „žinojimas” filosofijos, mokslo ir kasdienybės žaidimuose vartojamas skirtingai, bet šiuos vartojimus sieja šeiminiai panašumai – visi jie susiję su informacijos, patikrinimo ir užtikrintumo aspektais. Kitaip tariant, šeimyniniai panašumai rodo, kad kalbos vienybė nėra loginė ar metafizinė, bet praktinė – ji kyla iš mūsų gebėjimo atpažinti ir dalyvavti panašiose veiklos formose, kurios formuoja gyvą kalbos sistemą.

# LLM veikimo principai:

## Mokymosi duomenys.

## Dėmesio mechanizmas.

## Semantinių ryšių modeliavimas.

## GPT vs Bert? Tiesiog parodymui

# Literatūra

1. Wittgenstein, L. (1995). Tractaus Logico-Philosophicus, Filosofiniai tyrinėjimai. Rinktiniai raštai. Mintis: Vilnius.
2. Glock, H.J. (1996). A Wittgenstein Dictionary. Cambridge, Mass., USA: Wiley-Blackwell.
3. Schroeder, S (2006). Wittgenstein: The Way Out of the Fly-Bottle