

Kui hästi seletavad suured keelemudelid leksikograafide hinnangul eesti sõnade tähendusi?

Maria Tuulik, Ene Vainik, Esta Prangel, Margit Langemets, Eleri
Aedmaa, Kristina Koppel, Lydia Risberg

Teadusprojekt EKKD-III1 „Suurte keelemudelite rakendamine
leksikograafias: uued võimalused ja väljakutsed

Mida tahtsime teada?

- Kas ja mil määral saavad suured keelemudelid (SKMid) leksikograafi aidata sõnastikuseletuste loomisel ja sõnatähendus(t)e esitusel?
- Kuidas hindavad leksikograafid kui kvaliteedi kontrollijad ja tagajad SKM-ide genereeritud sisu?

Selle selgitamiseks viisime läbi katse, kus lasime SKM-idel seletada eesti keele sõnu ja seejärel tulemust hinnata leksikograafidel.

KATSE

Katsesõnad

- Valimis oli 60 sisusõna (võrdselt nimisõnad, omadussõnad, tegusõnad ja määrsõnad)
- Igast sõnaliigist 5 ühe tähendusega, 5 kahe tähendusega, 3 kolme tähendusega ja 2 nelja tähendusega.
- Katses esindatud sõnad valisime juhumeetodil EKI ühendsõnastikust (ÜS).
- Võõrsõnu sattus valimisse 15 ning omasõnu 45.

SKM-ide eelvalik

- o Llama 3
- o GPT-4
- o GPT-3.5-turbo
- o **GPT-4o**
- o **Gemini 1.5 Pro**
- o Claude 3 Sonnet

Prompt (*viip*)

„Mida tähendab eesti keeles [...]?”

Mitmetähenduslikele sõnadele võid anda mitu tähendust. Kui sa ei tea, siis ütle, et sa ei tea.“

Seletuse näide: *asuursinine*

GPT:

Asuursinine tähendab väga heledat ja kirkalt sinist värvitooni, mis meenutab selget taevast.

Gemini:

See on erksinise värvi varjund, mis meenutab selget taevast või sügavat vett. Asuursinine on sageli seotud rahu, vaikuse ja avarusega.

Küsitlus

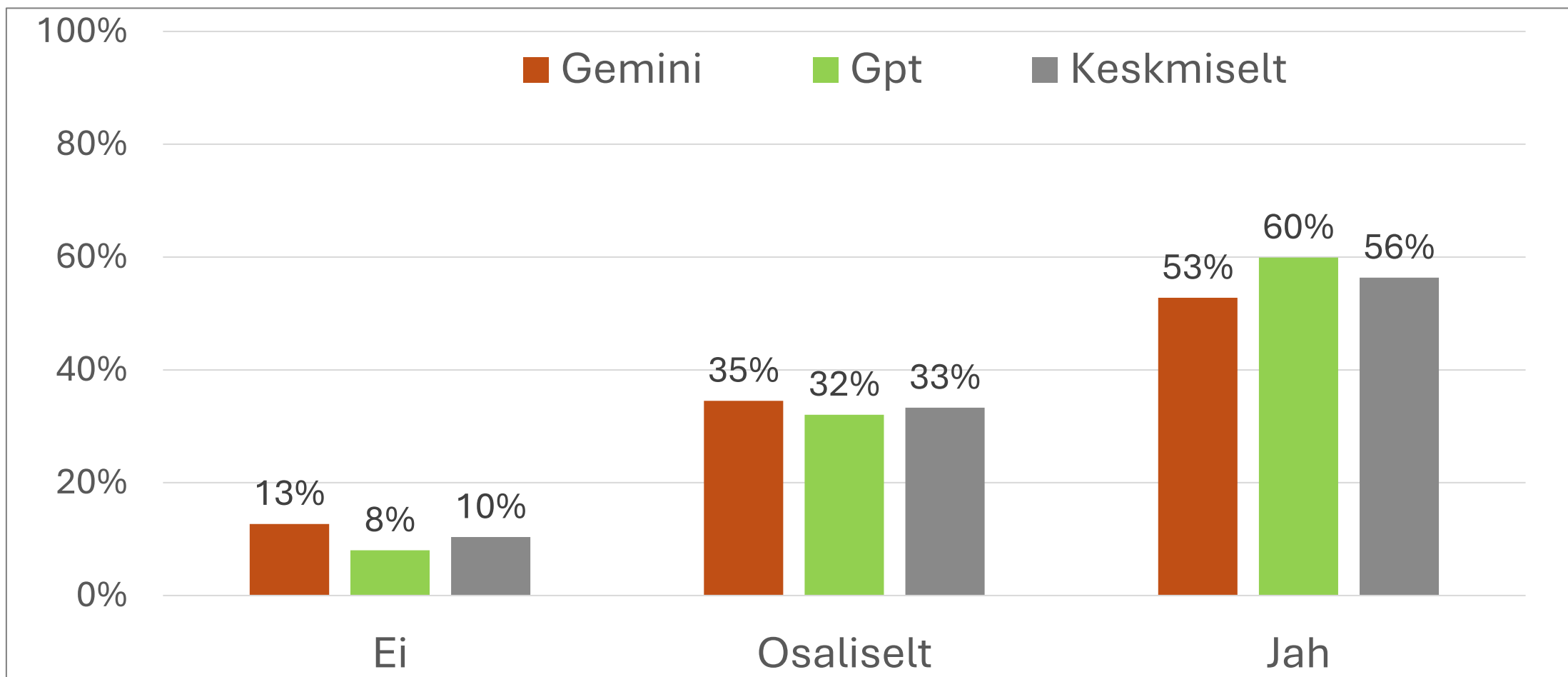
- 3 küsimust vastajate leksikograafilise tausta kohta:
 1. töökogemuse pikkus
 2. Sõnastikukogemuse laad
 3. seletusloome praktika
- 60 katsesõna - GPT ja Gemini seletused iga sõna kohta – kokku 120 hinnatavat stiimulit
- Kolm skalaarset küsimust mudelite soorituse kohta:
 1. tähenduste tabamine;
 2. vastuse keeleline korrektsus;
 3. pakutud seletusega rahulolu küsimus.
- Neljanda küsimuse all oli võimalik oma hinnangut vabas vormis kommenteerida.

Vastajad

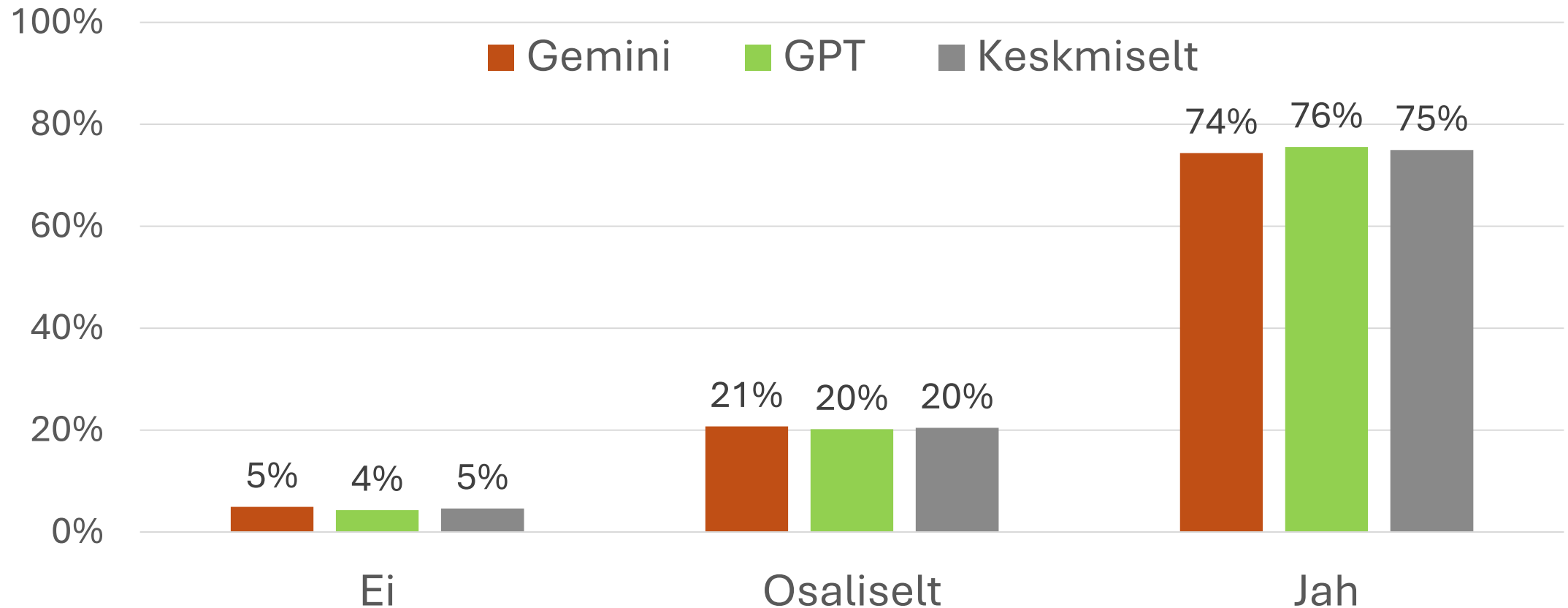
- 21 Eesti Keele Instituudi (EKI) leksikograafi, kelle töökogemus varieerus 3 aastast 43 aastani (keskmine 19; StDev 13,1)
- 14 vastajat oli töötanud ükskeelsete sõnastikega ning 7 vastajat terminibaasidega
- Pooltel vastajatest oli kogemus ise tähendusseletuste loomisega ja pooltel seletuste muudest allikatest otsimise (ja kohandamisega).

TULEMUSED

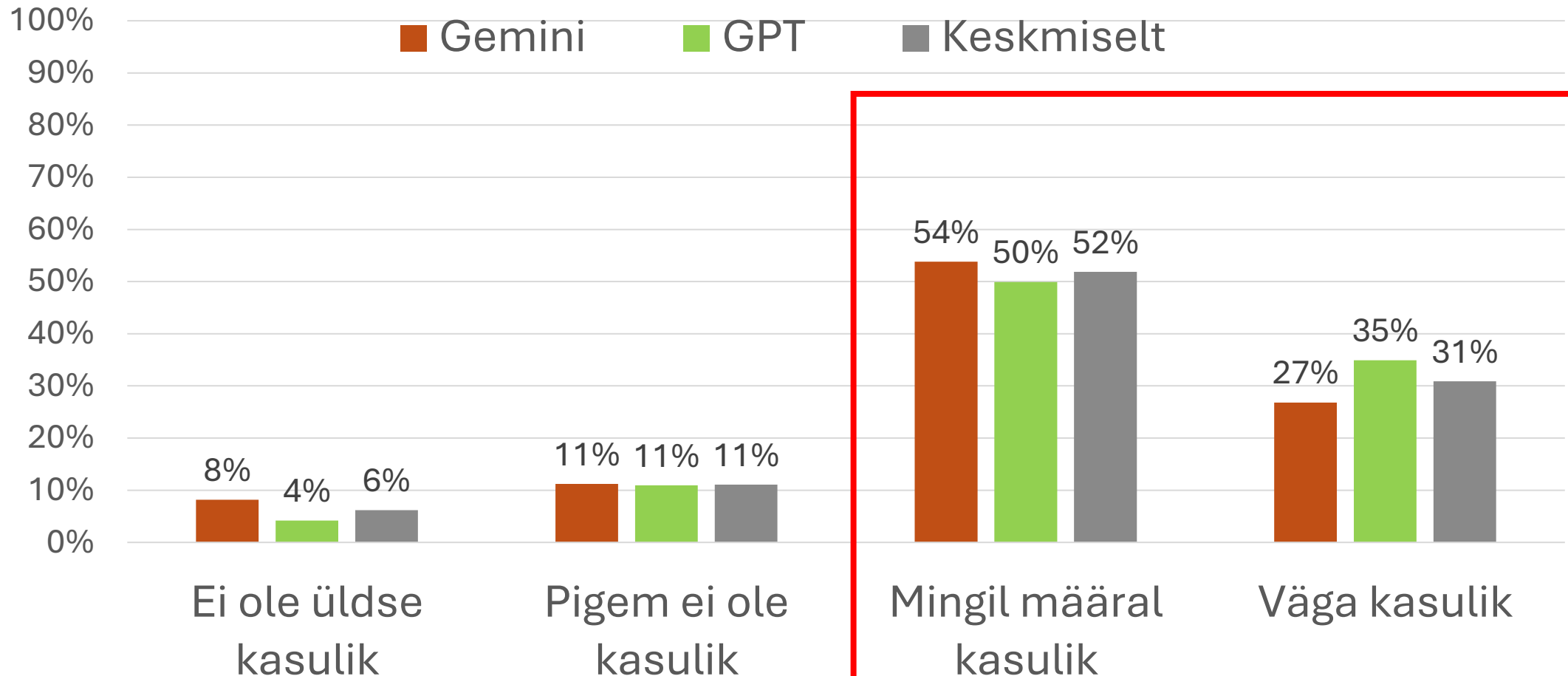
Tähenduste tabamine: „Kas antud tähendus(ed) on sinu arvates või sõnastiku põhjal õige(d)/olemasolev(ad) tähendus(ed)?“



Keeleline korrektsus: „Kas tähenduse seletus on grammatiliselt korrektses keeles?“



Kasulikkushinnang: „Kui kasulikuks hindad mudeli pakutud tähenduse seletust sinu kui leksikograafi jaoks?“



Lisaanalüüs

- Lisaparametrid: sõnaliik, tähenduste arv, mitmesõnalisus, korpussagedus, võõrsõnalisus
- Mitmesõnalisus oli negatiivselt seotud nii grammatilise korrektsuse hinnangu kui ka üldise kasulikkushinnanguga
- Võõrsõnalisus oli kasulikkushinnanguga nõrgas positiivses seoses
- Nõrk negatiivne seos leksikograafide staaži ja viljeldud seletusloome praktikaga: kogenumad ja tähendusesitusi ise loonud vastajad on hinnanud mudelite pakutud esitusi veidi kriitilisemalt
- Kasulikkushinnang ei olnud statistiliselt olulises vastastikuses seoses sõna polüseemsuse, sõnaliigi ega korpussagedusega

Vastajate eelistused erinesid (Fleissi kapp)

- Leksikograafide eelistused tähendusseletuse sisu ja vormi suhtes erinevad.
- Kolmest peamisest küsimusest (tähenduste tabamine, keeleline korrektsus ja kasulikkushinnang) varieerusid vastused enim (keskmine $\kappa = 0,12$) kasulikkushinnangu andmisel („Kui kasulikuks hindad mudeli pakutud tähenduse seletust sinu kui leksikograafi jaoks?“), mis oli ka kõige subjektiivsemat vastust eeldav küsimus.
- Täielikul üksmeelel ($\kappa = 1$) oldi ainult 8 tähendusseletuse keelelise korrektsuse hindamisel.

KOMMENTAARIDE KVALITATIIVNE ANALÜÜS

Tähenduste esitamisega seotud märkused

- Puuduvad tähendused
- Leiutatud tähendused
- Tähendused liiga üldised
- Arusaamatu tähendusjaotus polüseemse sõna juures
- Puuduvad homonüümid

Seletustega seotud märkused

- Liiga pikk (eriti GPT)
- Ebavajalik lisainfo kirjelduses
- Arusaamatu
- Kohmakas sõnastus
- Liiga napp (eriti Gemini)
- Seletus viitab valele sõnaliigile
- Küsitava sõna enda või selle tüve kasutus seletuses
- Seletatud ainult sünonüümide kaudu
- Liigselt terminoloogiline
- Ei võta arvesse konkreetse koha eripära

Näidetega seotud märkused

- Sisuliselt kahtlustäratavad või valed, nt **kostitama** 'kellelegi midagi osaks saada laskma' - *Ta kostitas teda kõvasti näkku.*
- Grammatiliselt ebakorrektseid
- Katsesõna puudub näitest
- Tähendus ja näide ei lange kokku
- Ebaloomulikud, nt *saledad puutüved, salk lilli*

Ülevaade positiivsetest märkustest

- Seletuste põhjalikkus, detailsus ja hea sõnastust
- Tähenduste ammendav kirjeldus
- Välja pakutud tähendused (ja näited), mida sõnastikus esitatud ei olnud (võib olla nii tugevus kui nõrkus)
- Head näitelaused
- Registriinfo esitamine

KOKKUVÕTE

Kokkuvõte 1

- Üldine rahulolu oli suhteliselt kõrge: leksikograafid hindasid oma töö jaoks väga kasulikuks või mingil määral kasulikuks 83% tähenduskirjeldustest.
- Mudelite sooritus eesti keeles enamasti rahuldas vastajaid: 75% juhtudel hinnati kasutatud keelt grammatiliselt korrektseks, 20% osaliselt korrektseks ja ainult 5% seletuste puhul täiesti ebakorrektseks.
- Rahulolu semantilise esitusega oli mõnevõrra madalam: väljatoodud tähendusi peeti õigeks 56% juhtudel, osaliselt õigeks 33% juhtudel ja valeks 10% juhtudel.
- GPT-4o ületas eesti keele sõnatähenduste kirjeldamisel Gemini 1.5 Pro-d

Kokkuvõte 2

- Keskmiselt hinnati võõrsõnadele antud seletusi kasulikumaks ja õigemaks
- Hinnanguid ei mõjutanud sõna tähenduste arv, sõnaliik ega korpussagedus
- Leksikograafide ootused heale seletusele erinevad. Tulevikus saame kasutada kasutada väljaselgitatud ootusi üksikasjalike promptide loomiseks, näiteks kohandada seletuse pikkust, sõnastuse keerukust, vältida tüve või märksõna ennast, vältida vormihomonüümiat (andes ette, et tegemist on lemmaga), vältida üksnes sünonüümidest koosnevaid definitsioone ja lisada või välistada näiteid.
- Praegu saavad SKM-id pakkuda põhja, mida tuleb leksikograafil kontrollida ja kohandada, kuid mis saab juba märkimisväärselt abiks olla nii tähenduste esitamisel/jagamisel kui ka seletuste sõnastamisel.

Lisamaterjal

- Võrreldes ÜSi tähenduste esitusega, kaldusid SKMid tähendusi pigem koondama.

