

Eesti keel suurtes keelemudelites

ETA küberturvalisuse komisjoni konverents, 22.09.2025

Kairit Sirts

TÜ arvutiteaduse instituut, keeletehnoloogia kaasprofessor

Eesti Tehisintellekti Tippkeskuse EXAI rühmajuht

Eesti keel suurtes keelemudelites

 **GPT, GPT-2, BERT**
esimesed
inglisekeelsed
mudelid

2018

 **GPT-3**
Esimene suur mudel
(nn LLM). Peamiselt
inglisekeelne

2020

2019-2020

 **mBERT, XLM-RoBERTa, EstBERT**
mitmekeelsed ja
keelespetsiifilised
mudelid

2022

 **ChatGPT**
Läbimurre,
mitmekeelne tugi, sh
eesti keel

 **GPT-4**
Parem eesti keel

 **LLaMa-2**
Avatud mudel,
ainult inglisekeelne

2023

 **LLaMa-3**
Mitmekeelsus,
sh eesti keel

2024

Küsimused

1. Kui hea on eesti keele tase erinevates keelemudelites?
2. Kas me (eestlastena) peaksime midagi tegema eesti keele taseme parandamiseks keelemudelites?

Keelemudelite hindamise meetodid

1. **Inimese mulje** – subjektiivne “vibe check”: kas vastus tundub hea?
2. **Automaatne testimine** – testandmestikud, mille tulemusi saab automaatselt skoorida

Vali lünka _ sobiv sõna:

Kuna pesumasin oli katki, viis Heiki Peetri riided pesumajja, sest _ pidi sel päeval tööl viibima.

Valik 1: Heiki

Valik 2: Peeter

Valik 2: Peeter

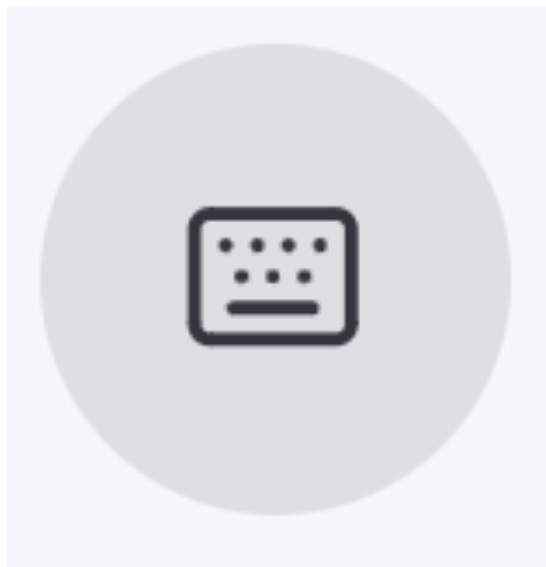
Põhjus: Lauses öeldakse, et Peeter viis riided pesumajja, seega on loogiline, et Peeter pidi sel päeval tööl viibima.

- Näide andmestikust: [winogrande_et](#)

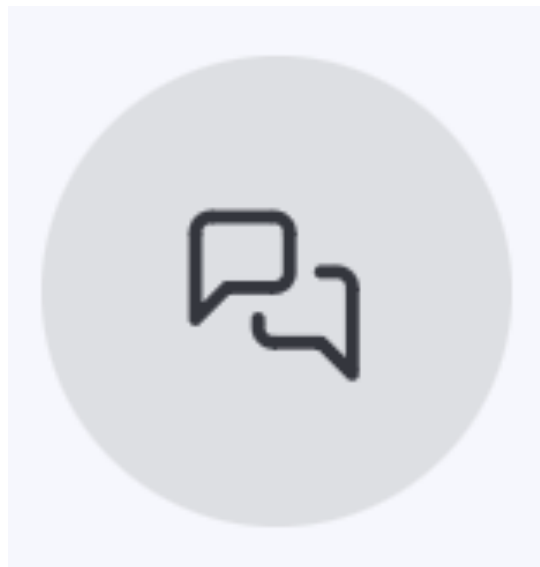
Keelemudelite hindamise meetodid

1. **Inimese mulje** – subjektiivne “vibe check”: kas vastus tundub hea?
2. **Automaatne testimine** – testandmestikud, mille tulemusi saab automaatselt skoorida
3. **Rahvahinnang** – inimeste hääletus eri mudelite vastuste vahel (Chatbot Arena, Tehisaru baromeeter)

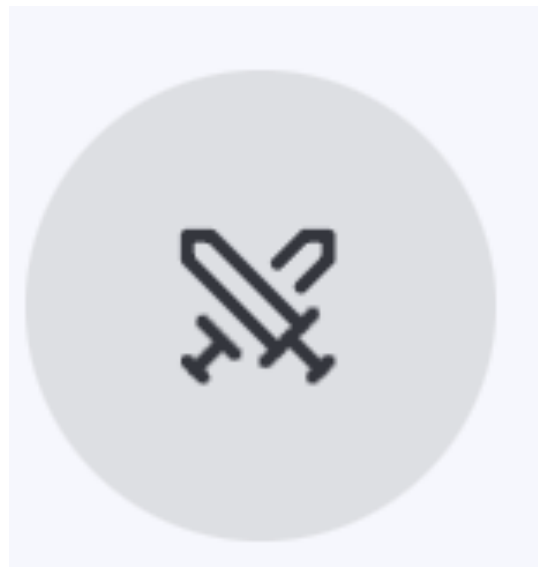
**1. Sisesta küsimus /
ülesanne**



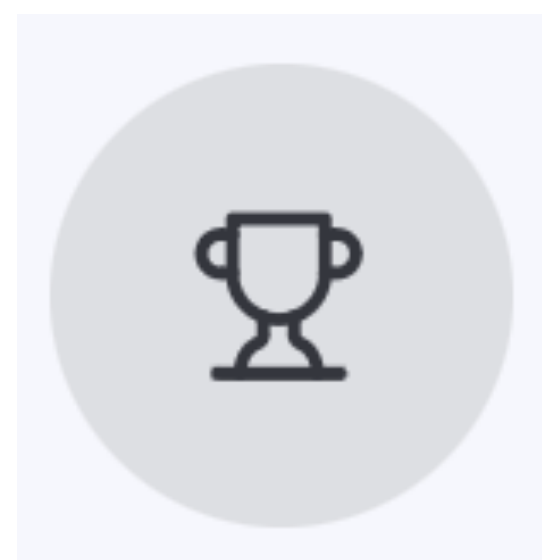
**2. Kaks juhuslikult
valitud mudelit vastavad**



**3. Vali kahest vastusest
parem**



**4. Vaata, mis mudelid
need olid**



Text Arena

View rankings across various LLMs on their versatility, linguistic precision, and cultural context across text.

Last Updated
Sep 8, 2025

Total Votes
4,075,191

Total Models
239

🏆 Overall

Search by model name...

/

Default

Rank (UB) ↑	Model ↑↓	Score ↑↓	95% CI (±) ↑↓	Votes ↑↓	Organization ↑↓	License ↑↓
1	 gemini-2.5-pro	1455	±5	41,731	Google	Proprietary
1	 claude-opus-4-1-20250805-thinking-16k	1451	±6	11,750	Anthropic	Proprietary
2	 o3-2025-04-16	1444	±4	43,898	OpenAI	Proprietary
2	 gpt-5-high	1442	±6	15,076	OpenAI	Proprietary
2	 chatgpt-4o-latest-20250326	1441	±4	36,426	OpenAI	Proprietary
3	 gpt-4.5-preview-2025-02-27	1439	±6	15,271	OpenAI	Proprietary
3	 claude-opus-4-1-20250805	1438	±6	18,341	Anthropic	Proprietary
5	 gpt-5-chat	1430	±6	11,808	OpenAI	Proprietary
6	 qwen3-max-preview	1428	±7	8,781	Alibaba	Proprietary
8	 grok-4-0709	1422	±5	21,446	xAI	Proprietary

View all

Tehisaru baromeeter: baromeeter.ai

Eestikeelne Chatbot Arena



Vestle



Edetabel



Meist



Kasutustingimused

🇪🇪 TEHISARU BAROMEETER 🇪🇪

1. Küsi eesti keeles
2. Vali parim vastus
3. Tutvu edetabeliga!



Kliki siia, et näha võrdluses olevaid mudeleid. 🔥 Valikus uued mudelid!

💬 Mudel A

💬 Mudel B

👉 Kirjuta siia enda küsimus ja vajuta ENTER

Saada

Vali lünka _ sobiv sõna:

Kuna pesumasin oli katki, viis Heiki Peetri riided pesumajja, sest _ pidi sel päeval tööl viibima.

Valik 1: Heiki

Valik 2: Peeter

Valik 2: Peeter

Selgitus: Lause kontekstist lähtudes on loogiline, et Peeter pidi sel päeval tööl viibima ja seetõttu viis Heiki tema riided pesumajja. Kui Heiki ise oleks pidanud tööl olema, poleks tal olnud võimalik Peetri riideid pesumajja viia.

A on parem



Viik



Kirjuta siia enda küsimus ja vajuta ENTER

Vali lünka _ sobiv sõna:

Kuna pesumasin oli katki, viis Heiki Peetri riided pesumajja, sest _ pidi sel päeval tööl viibima.

Valik 1: Heiki

Valik 2: Peeter

Valik 2: Peeter

Põhjus: Lauses öeldakse, et Peeter viis riided pesumajja, seega on loogiline, et *Peeter* pidi tööl viibima. Heiki on lihtsalt tegevuse sooritanud isik.

B on parem



Mõlemad on halvad

Saada

Vali lünka _ sobiv sõna:

Kuna pesumasin oli katki, viis Heiki Peetri riided pesumajja, sest _ pidi sel päeval tööl viibima.

Valik 1: Heiki

Valik 2: Peeter

Vali lünka _ sobiv sõna:

Kuna pesumasin oli katki, viis Heiki Peetri riided pesumajja, sest _ pidi sel päeval tööl viibima.

Valik 1: Heiki

Valik 2: Peeter

Valik 2: Peeter

Selgitus: Lause kontekstist lähtudes on loogiline, et Peeter pidi sel päeval tööl viibima ja seetõttu viis Heiki tema riided pesumajja. Kui Heiki ise oleks pidanud tööl olema, poleks tal olnud võimalik Peetri riideid pesumajja viia.

Valik 2: Peeter

Põhjus: Lauses öeldakse, et Peeter viis riided pesumajja, seega on loogiline, et *Peeter* pidi tööl viibima. Heiki on lihtsalt tegevuse sooritanud isik.

Mudel A: claude-3-opus-20240229

Mudel B: gemma-3-12b-it

A on parem



Viik

B on parem



Mõlemad on halvad

Kliki "🗨️ Uus vestlus" et uut vestlust alustada.

Saada

Edetabel

Tabel

Mudeleid kokku: 48. Hääli kokku: 12 367 / 50 000. Tabelit uuendatakse regulaarselt. Viimati uuendatud: 2025-09-15 15:51:14.

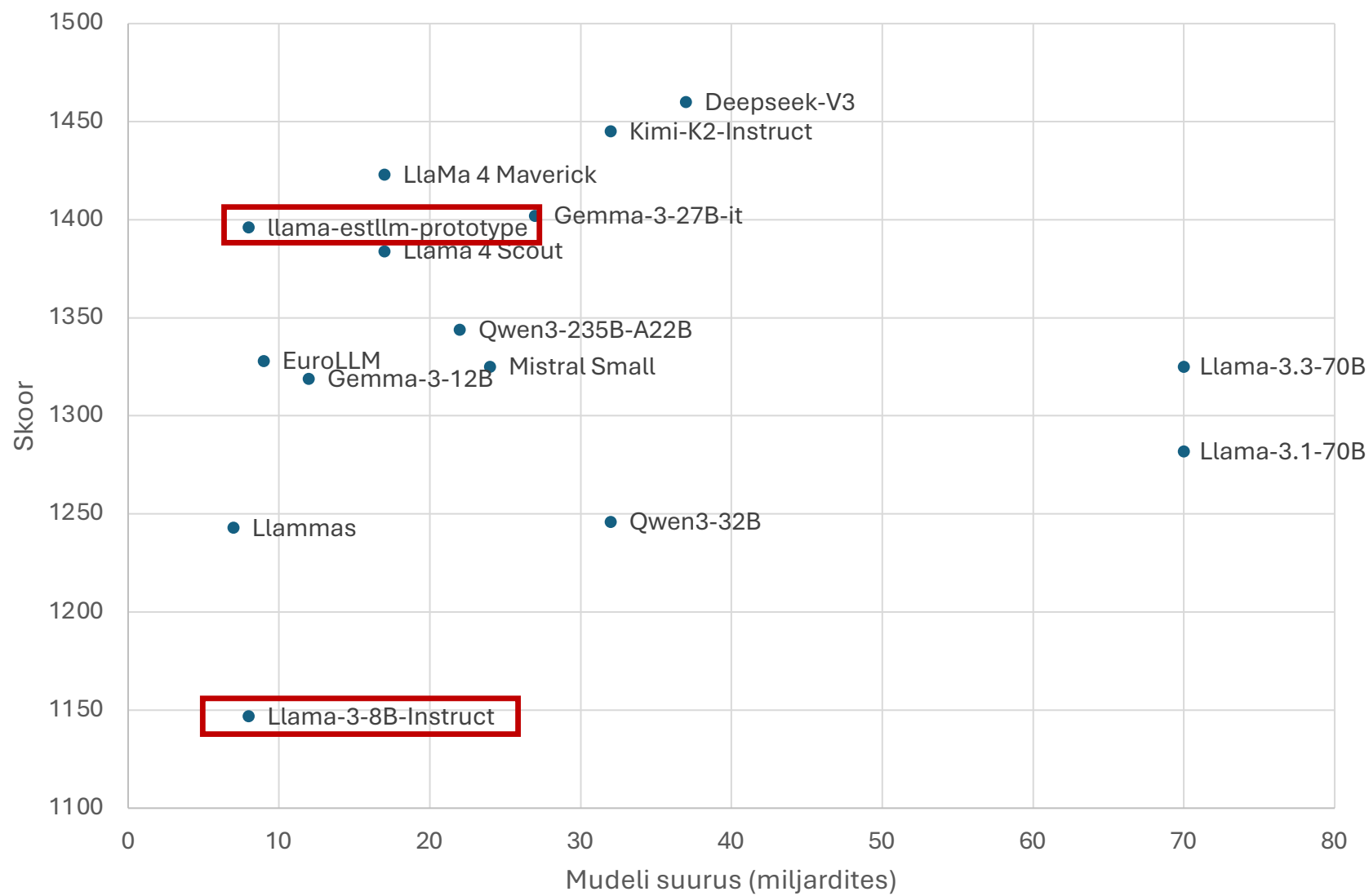
☐ ainult avalikud mudelid

☒ peida vanad versioonid

Koht	Mudel	Skoor	95% UI	Hääli	Looja	Litsents	Teadmiste lõpp
1	Gemini-2.5-Pro-Exp-06-05	1639	+52/-41	350	Google	Proprietary	Unknown
2	GPT-4.1-2025-04-14	1535	+44/-41	459	OpenAI	Proprietary	Unknown
2	Grok 3 (beta)	1530	+39/-33	926	xAI	Proprietary	Unknown
2	Gemini-2.5-Flash-Exp-04-17	1517	+46/-29	516	Google	Proprietary	Unknown
2	GPT-4o-2024-11-20	1516	+38/-41	441	OpenAI	Proprietary	Unknown
2	Claude-4 Opus (20250514)	1509	+37/-28	459	Anthropic	Proprietary	Unknown
2	Claude-3.7 Sonnet (20250219)	1494	+41/-29	982	Anthropic	Proprietary	Unknown
2	Claude 3.5 Sonnet (20241022)	1491	+36/-25	945	Anthropic	Proprietary	2024/4
2	Gemini-2.0-Flash-001	1489	+31/-26	966	Google	Proprietary	2024/08
2	Claude 3 Opus (20240229)	1475	+34/-32	999	Anthropic	Proprietary	2023/8

☒ ainult avalikud mudelid☒ peida vanad versioonid

Koht ▲	Mudel ▲	Skoor ▲	95% UI ▲	Hääli ▲	Looja ▲	Litsents ▲	Teadmiste lõpp ▲
1	DeepSeek-V3 (0324)	1460	+33/-33	637	DeepSeek	MIT	Unknown
1	Kimi-K2-Instruct	1445	+39/-38	386	Moonshot AI	Modified MIT	Unknown
1	Llama 4 Maverick	1423	+37/-34	392	Meta	Apache 2.0	Unknown
1	Gemma-3-27B-it	1402	+33/-35	446	Google	Gemma	Unknown
1	llama-estllm-prototype-0825	1396	+49/-44	177	TartuNLP	Llama 3.1	Unknown
2	Llama 4 Scout	1384	+36/-33	410	Meta	Apache 2.0	Unknown
3	Meta-Llama-3.1-405B-Instruct	1365	+36/-36	449	Meta	Llama 3.1 Community	2023/12
4	Qwen3-235B-A22B	1344	+36/-31	430	Alibaba	Apache 2.0	Unknown
4	EuroLLM-9B-Instruct	1328	+42/-33	437	UTTER	Apache 2.0	Unknown
4	Gemma-3-12B-it	1319	+50/-53	155	Google	Gemma	Unknown
6	Llama-3.3-70B-Instruct	1325	+27/-35	718	Meta	Llama 3.3 Community	Unknown
8	Meta-Llama-3.1-70B-Instruct	1282	+38/-47	230	Meta	Llama 3.1 Community	2023/12
9	Mistral Small 3.1 (2503)	1269	+36/-31	629	Mistral	Apache 2.0	Unknown
11	Qwen3-32B	1246	+41/-33	467	Alibaba	Apache 2.0	Unknown
11	Llammas	1243	+45/-32	406	TartuNLP	Llama 2	Unknown



Keelemudelite hindamise meetodid

1. **Inimese mulje** – subjektiivne “vibe check”: kas vastus tundub hea?
2. **Automaatne testimine** – testandmestikud, mille tulemusi saab automaatselt skoorida
3. **Rahvahinnang** – inimeste hääletus eri mudelite vastuste vahel (Chatbot Arena, Tehisaru baromeeter)
4. **Mudelite omavaheline hindamine** – üks LLM hindab teise LLM'i vastuseid (“LLM as a judge”)

Miks me peaksime ise keelemudelite arendamisega tegelema?

- **Andmed:** potentsiaalne võimalus kaasata rohkem andmeid
- **Kompetents:** oskus mudeleid treenida ja täiendada
- **Avatud mudelid:** saab kasutada tundlike ja konfidentsiaalsete andmete töötlemiseks





Mida ei ole mõistlik teha?

- **Nullist “Eesti GPT”** → pole realistlik, andmeid liiga vähe
- **Andmed suurtele firmadele** → ei saa kasutada kõike, eriti autoriõigustega sisu
- **Süntetilised või masintõlgitud andmed** → Ei asenda pärisandmeid

Millega tegeleb EstLLM projekt?



Andmed



Treenimine



Hindamine

Aitäh!