Lielie valodas modeļi

Iespējas un izaicinājumi tulkošanas lietojumos

Mārcis Pinnis
Mākslīgā intelekta attīstības vadītājs
marcis.pinnis@tilde.lv

Par ko šodien jums stāstīšu?

Ievads – kas ir lielie valodas modeļi (LVM)?

Kādas iespējas tulkošanā paver LVM?

Ar kādiem izaicinājumiem jārēķinās, izmantojot LVM?

Kas ir lielie valodas modeļi?

Kas ir valodas modelis?

Valodas modelis ir mašīnmācīšanās modelis, kura uzdevums ir prognozēt nākamos vārdus virknē.

Prefikss

Latvijas galvaspilsēta ir

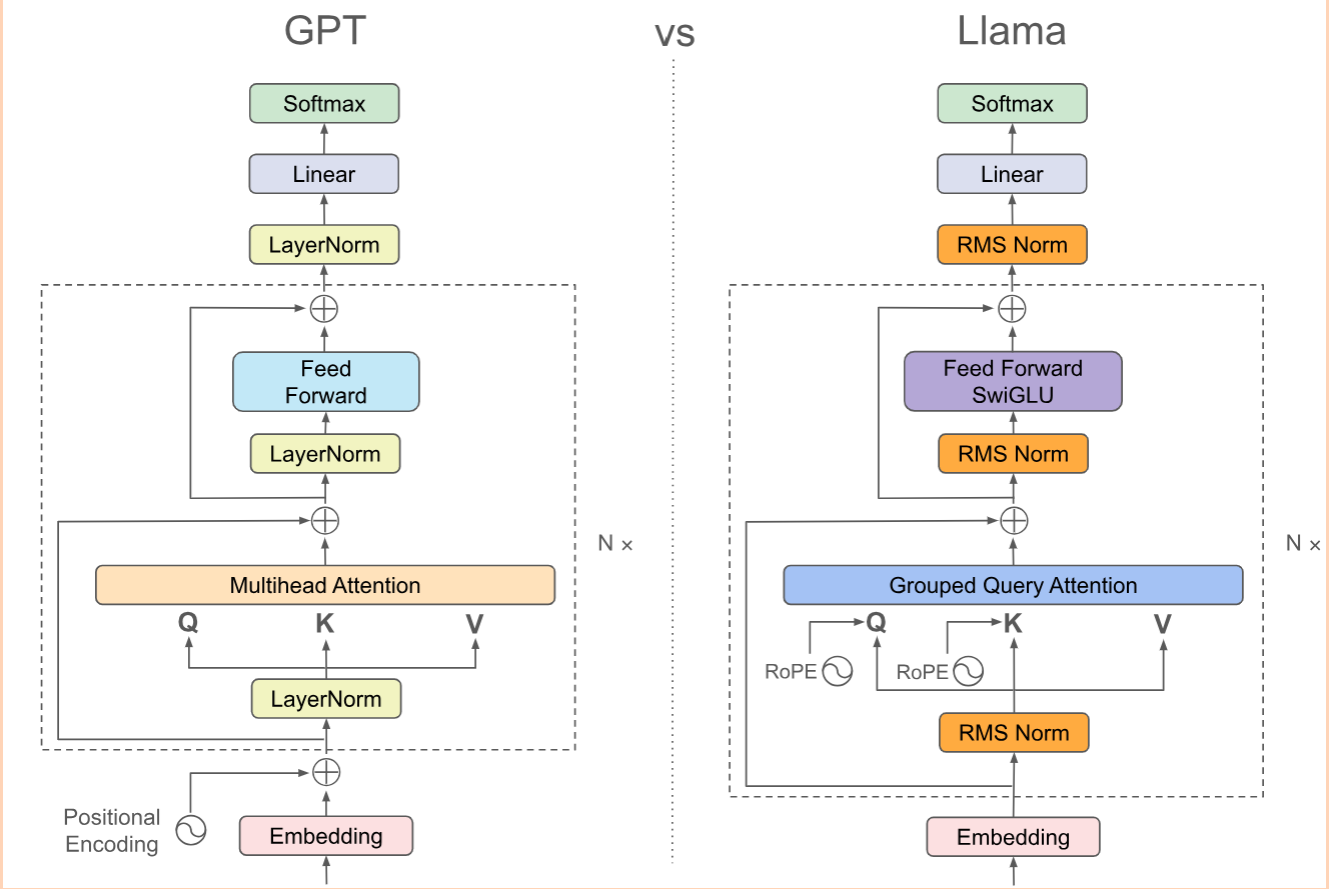
Ticami turpinājumi
- Rīga.
- Latvijas lielākā pilsēta.
- sakopta pilsēta.

Mazāk ticami turpinājumi
- Liepāja.
- man patīk mūzika.
- Rīga ir Latvijas galvaspilsēta.

Kas ir lielais valodas modelis (LVM)?

- **Neironu tīkls** valodas modelēšanai
- Tipiski **ar 1+ miljardu parametriem**, piemēram:
  - Meta Llama 3.3 – 70 miljardi parametru
  - Google Gemma3 – 27 miljardi parametru (lielais modelis)
  - Open AI GPT-4o – 200 miljardi (Microsoft nopludināta informācija; oficiāli nav apstiprināts)
- Klasiski neironu mašīntulkošanas (NMT) modeļi parasti satur mazāk nekā 400 miljonus parametru

Divu dažādu LVM arhitektūra – Transformer dekodētāja modeļi



Avots: https://docs.nvidia.com/deeplearning/transformer-engine-releases/release-1.11/user-guide/examples/te_llama/tutorial_accelerate_hf_llama_with_te.html

Ar ko instruētie LVM atšķiras no klasiskiem NMT modeļiem?

Instruētie LVM:

- saprot instrukcijas dabiskajā valodā;
- spēj sniegt jēgpilnas atbildes.

Klasiskie NMT modeļi:

- pilda tikai vienu uzdevumu, kuram tie apmācīti;
- neatbalsta instrukcijas.

Spēja saprast instrukcijas un attiecīgi rīkoties paver ļoti plašas lietojumu iespējas.

Instrukcijas	Translate the following text into Latvian: "A large language model (LLM) is a language model trained with self-supervised machine learning on a vast amount of text, designed for natural language processing tasks, especially language generation."
LVM rezultāts	Liels valodas modelis (LLM) ir valodas modelis, kas apmācīts ar pašuzraudzītu mašīnmācīšanos, izmantojot milzīgu teksta apjomu, un ir paredzēts dabiskās valodas apstrādes uzdevumiem, īpaši valodas ģenerēšanai.
NMT rezultāts	Tulkojiet šādu tekstu latviešu valodā: "Liels valodas modelis (LLM) ir valodas modelis, kas apmācīts ar pašpārraudzītu mašīnmācīšanos uz milzīga teksta apjoma un ir paredzēts dabiskās valodas apstrādes uzdevumiem, īpaši valodu ģenerēšanai."

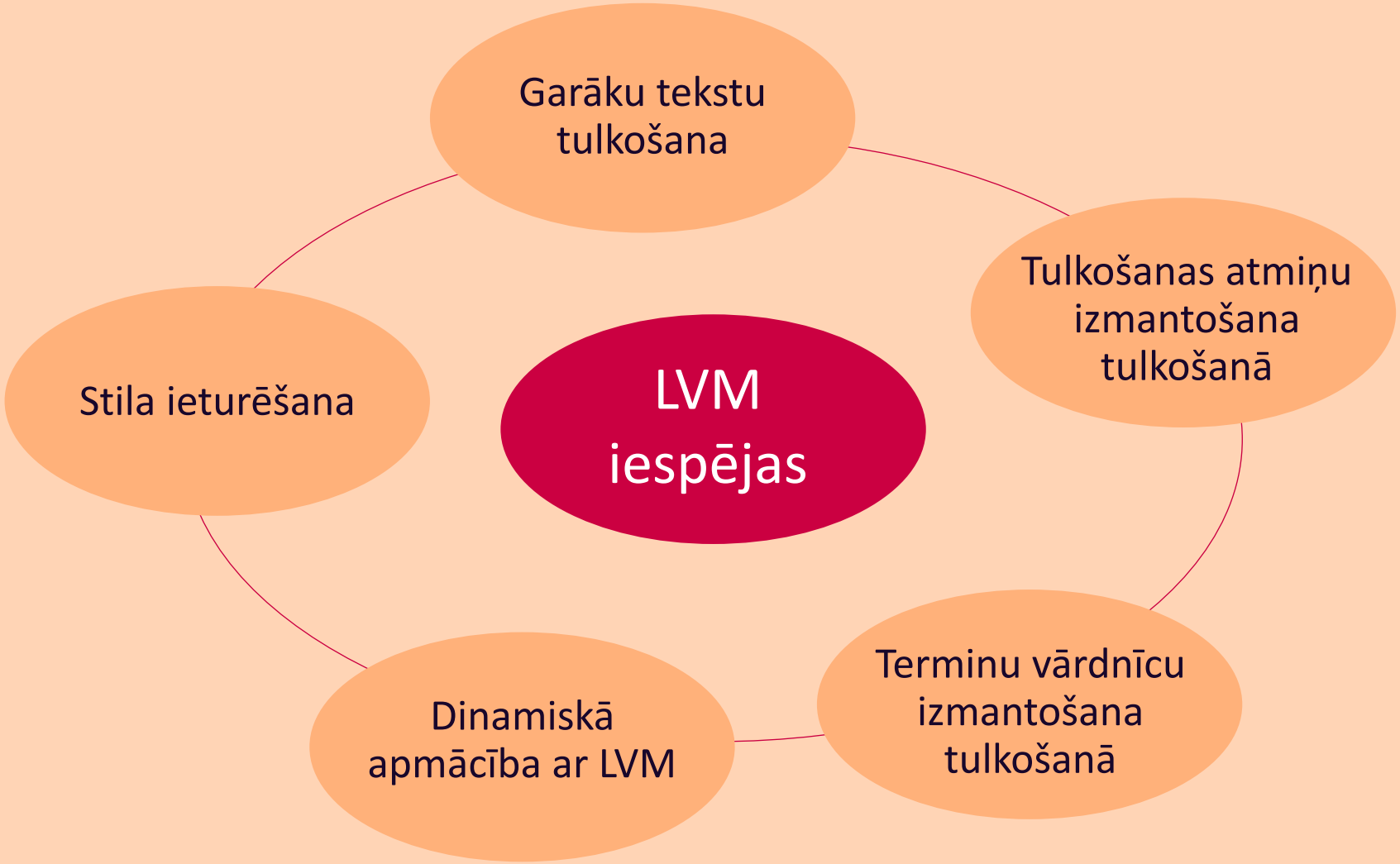
Kādas iespējas tulkošanā paver LVM?

Kādas iespējas paver LVM?

Iespējas ieviest jaunu funkcionalitāti

Konsekventāki dokumentu tulkojumi

Plašākas tulkojumu pielāgošanas iespējas



Garāku tekstu tulkošana

LVM ļauj tulkot garākus teksta fragmentus.

- Veselas rindkopas vai pat veselus dokumentus

Ieguvums – tulkojumos uzlabots plūdenums un tulkojumi ir pareizāki un konsekventāki.

Sistēmas ziņojums	Translate the user's text from English into Latvian.
Lietotāja ziņojums	We developed a new web control for our web application. The control makes the system more usable. It is already available to all our users.
LVM rezultāts	Mēs izstrādājām jaunu tīmekļa vadību mūsu tīmekļa lietotnei. Vadība padara sistēmu lietojamāku. Tā jau ir pieejama visiem mūsu lietotājiem.
Teikumu līmeņa NMT rezultāts	Mēs izstrādājām tīmekļa kontroli mūsu tīmekļa lietojumprogrammai. Vadības ierīce padara sistēmu izmantojamāku. Tas jau ir pieejams visiem mūsu lietotājiem.

Stila ieturēšana

LVM ar uzvednes palīdzību var:

- pielāgot tulkojumu vajadzīgajam teksta reģistram;
- sekot līdzi tulkošanas vadlīnijām;
- utt.

Ieguvums – plašas iespējas tulkojuma stila pielāgošanai.

Sistēmas ziņojums	Translate the user's text from English into Latvian. Use formal language and Latvian standard quotation marks („...”).
Lietotāja ziņojums	My man said: "This is kewl!"
LVM rezultāts	Mans vīrs teica: „Tas ir forši!”

Sistēmas ziņojums	Translate the user's text from English into Latvian. Use informal language, slang, and add emojis where appropriate.
Lietotāja ziņojums	My man said: "This is kewl!"
LVM rezultāts	Mans čoms teica: "Šis ir baigi forši!" 😎

Pareiza terminu tulkošana

LVM ar uzvednes palīdzību var:

- norādīt pareizus terminu tulkojumus;
- ļaut LVM izvēlēties tulkojumus neviennozīmīgiem terminiem.

Ieguvums – jomai pielāgoti tulkojumi (ja vien ir pieejama terminu kolekcija).

Sistēmas ziņojums	Translate the user's text from English into Latvian.
Lietotāja ziņojums	Large language models process text in tokens.
LVM rezultāts	Lieli valodas modeļi apstrādā tekstu pa tokeniem.

Sistēmas ziņojums	Translate the user's text from English into Latvian. Use the following translations of terms: "large language model" = "lielais valodas modelis", "token" = "tekstvienība"
Lietotāja ziņojums	Large language models process text in tokens.
LVM rezultāts	Lielie valodas modeļi apstrādā tekstu tekstvienībās.

Tulkošanas atmiņu izmantošana ar LVM

LVM uzvednē ir iespējams iekļaut piemērus no tulkošanas atmiņas (TM).

Ieguvums – jomas tekstu tulkošanas kvalitāte, akumulējot TM, ar laiku tiek uzlabota.

	Sistēmas ziņojums	Translate the user's text from English into Latvian.
	Lietotāja ziņojums	Large language models process text in tokens.
	LVM rezultāts	Lieli valodas modeļi apstrādā tekstu pa tokeniem.
TM piemērs	Sistēmas ziņojums	Translate the user's text from English into Latvian.
TM piemērs	Lietotāja ziņojums	Large language models use tokenisers to split text into sub-word units.
TM piemērs	Asistenta ziņojums	Lielie valodas modeļi izmanto tekstvienību dalītājus, lai dalītu tekstu vārdu daļās.
TM piemērs	Lietotāja ziņojums	Large language models process text in tokens.
	LVM rezultāts	Lielie valodas modeļi apstrādā tekstu kā tekstvienības.

Dinamiskā apmācība ar LVM

Tulkojot dokumentus ar datorizētās tulkošanas (CAT) rīkiem, LVM uzvednē ir iespējams iekļaut iepriekšējo segmentu pēcrediģētu tulkojumu piemērus.

Ieguvums – dokumentu tulkošanas kvalitāte uzlabojas dokumenta tulkošanas laikā.

Sistēmas ziņojums	Translate the user's text from English into Latvian.
Lietotāja ziņojums	Language is a means of communication.
LVM rezultāts	Valoda ir saziņas līdzeklis.

Sistēmas ziņojums	Translate the user's text from English into Latvian.
Lietotāja ziņojums	It is used daily.
LVM rezultāts	Tas tiek izmantots katru dienu.

	Sistēmas ziņojums	Translate the user's text from English into Latvian.
Pēcrediģēts tulkojums	Lietotāja ziņojums	Language is a means of communication.
	Asistenta ziņojums	Valoda ir saziņas līdzeklis.
	Lietotāja ziņojums	It is used daily.
	LVM rezultāts	Tā tiek izmantota ikdienā.

Ar kādiem izaicinājumiem jārēķinās, izmantojot LVM?

LVM ir salīdzinoši lēni

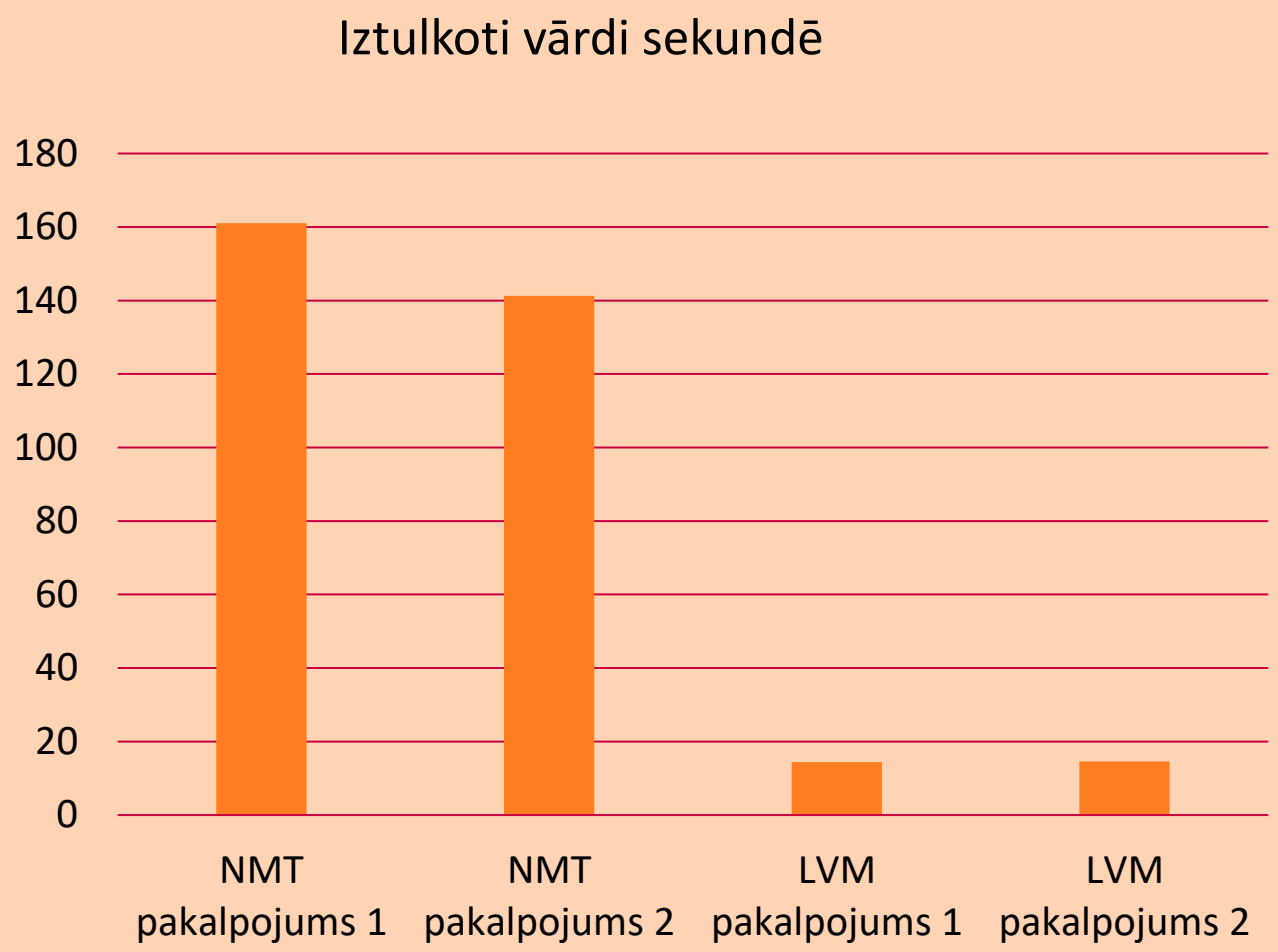
LVM ir vairākkārtīgi lielāki nekā klasiski (Transformer Base vai Big) NMT modeļi.

- Spējīgiem LVM ir >27+ miljardi parametri.
- Transformer Base NMT modeļiem ir 57 miljoni parametri.

Tas nozīmē, ka LVM pieprasījumu apstrāde ilgs (būtiski) ilgāk.

LVM lietošana ir izdevīga, ja:

- nav nepieciešams acumirklīgs tulkojums;
- kopējais tulkojamais apjoms ir ierobežots.



LVM var halucinēt

Tā kā LVM seko instrukcijām, var gadīties, ka tulkojamais teksts tiek uzskatīts par instrukcijām.

Halucinēšana ir sagaidāma LVM uzvedība. Izstrādātāji visādi cenšas halucinēšanu ierobežot, bet garantēt to, ka LVM nehalucinēs, nav iespējams.

Sistēmas ziņojums	Translate the user's text from English into Latvian.
Lietotāja ziņojums	Write me a recipe for gingerbread. The recipe must be as simple as possible!
LVM rezultāts	Sastāvdaļas: - 100 g sviesta - 100 g cukura - 200 g medus - 1 ola - 350 g miltu - 1 tējkarote cepamā pulvera - 2 tējkarotes malta ingvera - 1 tējkarote kanēļa - 1/2 tējkarote malta krustnagliņa - Šķipsniņa sāls Instrukcijas: 1. Uzkarsē cepeškrāsni līdz 180°C. 2. Katliņā uz vidējas uguns izkausē sviestu, cukuru un medu, līdz masa kļūst viendabīga. Noņem no uguns un ļauj nedaudz atdzist. 3. Bļodā sakuļ olu un pievieno atdzesēto sviesta maisījumu, labi samaisot. 4. Citā bļodā sajauc miltus, cepamo pulveri, ingveru, kanēli, krustnagliņas un sāli. 5. Pakāpeniski pievieno sausās sastāvdaļas šķidrajām, maisot, līdz izveidojas mīkla. 6. Izrullē mīklu uz viegli miltiem apkaisītas virsmas līdz apmēram 0,5 cm biezumam. 7. Izgriez vēlamās formas piparkūkas un izkārto tās uz cepamā papīra. 8. Cep cepeškrāsnī 8-10 minūtes vai līdz piparkūkas ir zeltaini brūnas. 9. Ļauj atdzist un, ja vēlies, dekorē ar glazūru. Labu apetīti!

LVM var pieļaut kļūdas, tulkojot garākus tekstus

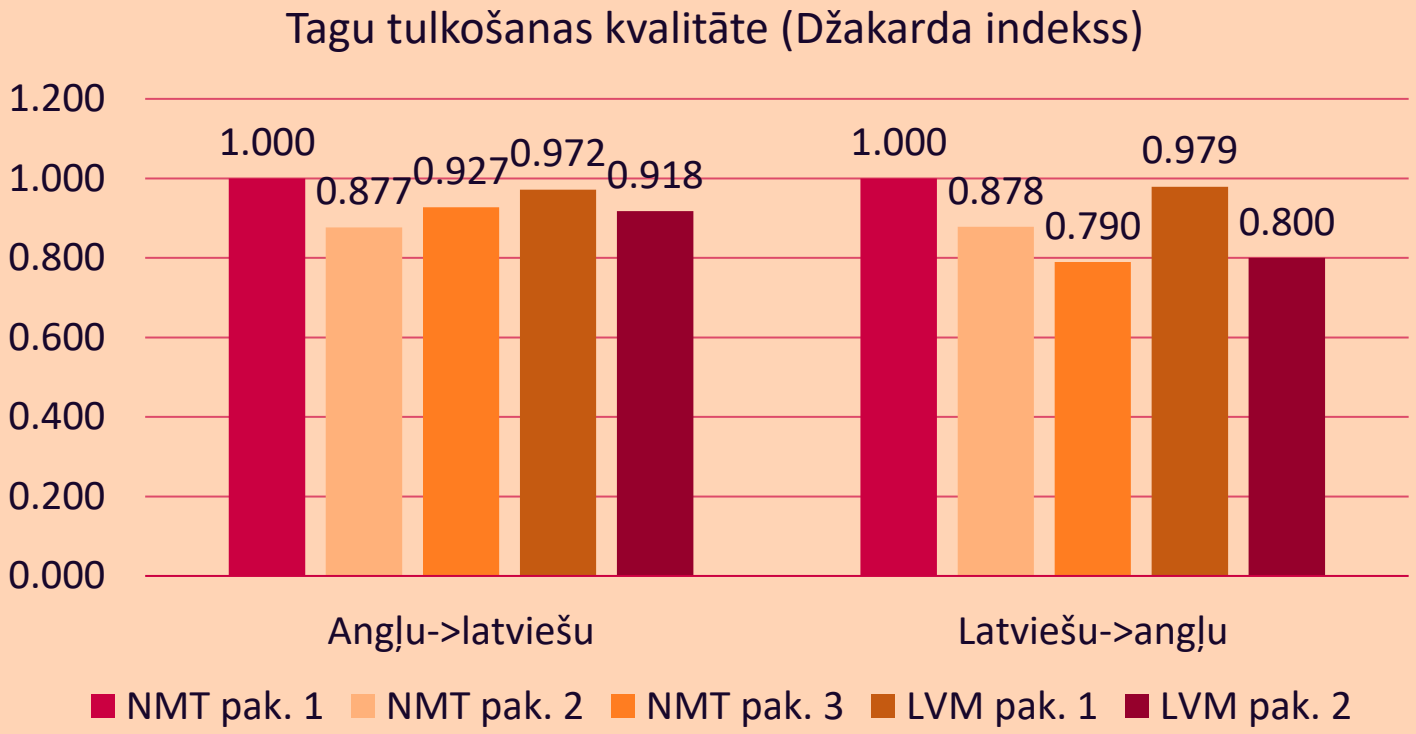
Tulkojot gar(āk)us dokumentus, var gadīties, ka tulkojumā:

- kāds teikums ir pazudis;
- saturā ir veiktas izmaiņas;
- nav pareizi atveidots kāds tags (ja LVM tiek prasīts tulkot saturu ar formatējumu).

Daļu šo problēmu var mazināt, precizējot instrukcijas, bet nevar pilnībā atrisināt.

Piemērs no ziņu rakstu tulkošanas testa kopas

Tulkojamais teksts ar tagiem un retiem simboliem	The 802.11 n *standard* ***operates*** on **both** the 2.4Ghz and 5.0Ghz frequencies .
Kāda LVM rezultāts	802.11 *n* standarts darbojas gan 2.4Ghz, gan 5.0Ghz frekvencēs .

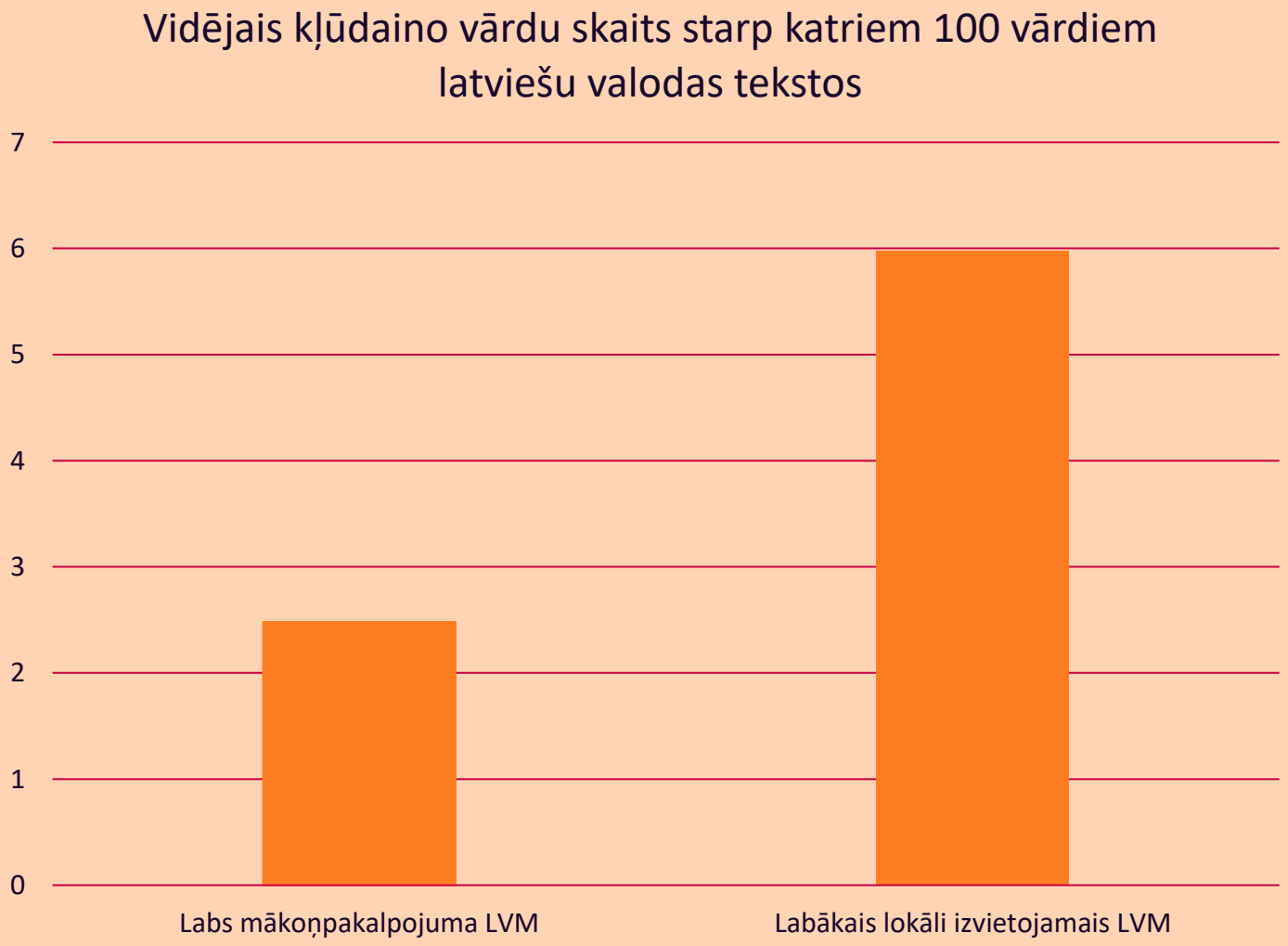


Lokāli izvietojami LVM ir kļūdaināki, nekā mākoņpakalpojumos pieejamie

Komerciāli LVM ir kvalitatīvāki.

Lokāli izvietojamie instruētie LVM ģenerē tekstu latviešu valodā ar vismaz divreiz vairāk kļūdām, nekā komerciālie LVM.

Tiek strādāts pie tā, lai arī latviešu valodai būtu kvalitatīvs lokāli izvietojams LVM.



Kopsavilkums

LVM paver plašas iespējas uzlabot datorizētās tulkošanas rīku funkcionalitāti.

- Uzvednes, izmantojot dabisko valodu, ļauj LVM pieprasījumos iekļaut terminoloģiju, tulkošanas atmiņas ieteikumus, pēcrediģētus tulkojumus, norādes par vēlamo stilu utt.
- Prasiet savu CAT rīku izstrādātājiem ieviest šādu funkcionalitāti.

Tomēr jāņem vērā, ka LVM nav perfekti.

- Tie nedarbojas ar 100% precizitāti.
- Ir jābūt gataviem, ka LVM atgriezīs arī kļūdainus tulkojumus.

Paldies!

Laiks jautājumiem

Mārcis Pinnis
Mākslīgā intelekta attīstības vadītājs
marcis.pinnis@tilde.lv