

# Slovenski ChatGPT se imenuje GaMS

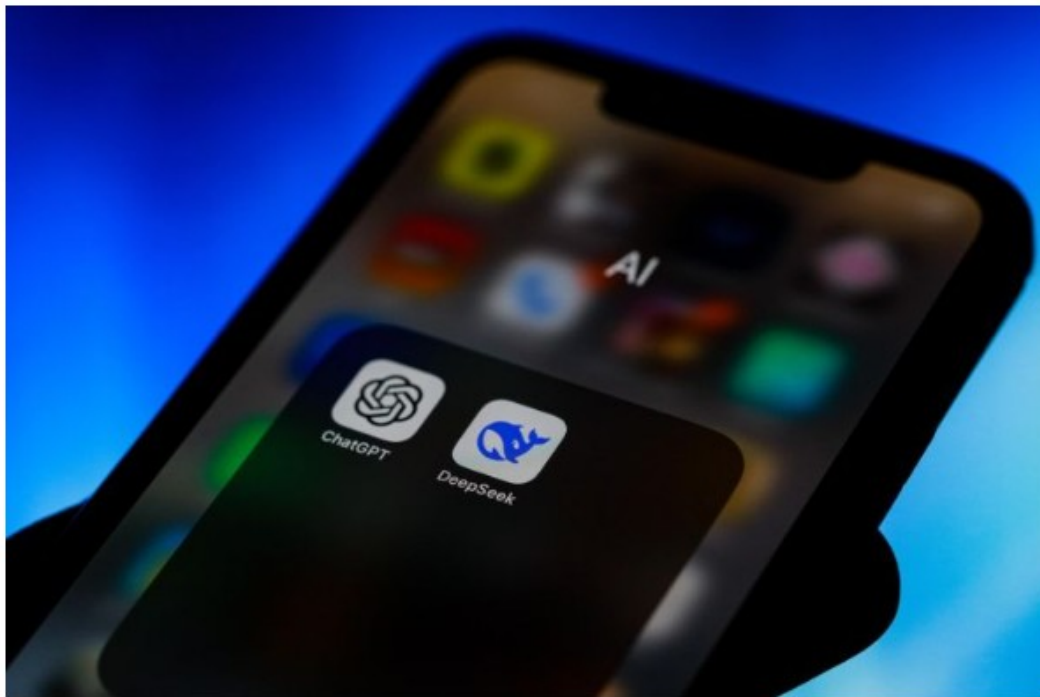
**Ljubljanska univerza** že razvija alternativo ChatGPT v slovenščini, ki si jo bo mogoče namestiti na osebni računalnik. Včeraj napovedan nov superračunalnik bo njen razvoj le še pospešil.



Avtor: Svet24.si / Luka Tetičkovič



sreda, 12. 3. 2025, 05:57



ChatGPT po kitajski konkurenci dobiva tudi evropsko, v sklopu katere se razvija tudi Generativni model slovenščine (GaMS). | Foto: Profimedia

V anketi Svet24 je na vprašanje, koliko si pri vsakdanjih opravilih pomagajo z umetno inteligenco, 19 odstotkov bralcev odgovorilo, da uporabljajo ameriški ChatGPT. Dva odstotka posegata po kitajskem DeepSeeku, trije odstotki pa po katerem od drugih velikih jezikovnih modelov. Devetindvajset odstotkov orodja ne uporablja, ker se jim ne zdi

koristno, skoraj polovica pa mu ne zaupa. Glasovalo je 875 oseb.

Komercialnim modelom ne zaupajo oziroma je njihova uporaba zaradi občutljivosti podatkov v državni upravi, sodstvu, zdravstvu in nekaterih podjetjih celo nemogoča, je za Svet24 povedal profesor ljubljanske **Fakultete za računalništvo in informatiko** (FRI UL) **Marko Robnik Šikonja**.

FRI UL v sodelovanju s Centrom za jezikovne vire in tehnologije **Univerze v Ljubljani** v projektu PoVeJMo razvija slovenski jezikovni model, ki bo brezplačen in ga bo mogoče uporabljati tudi lokalno – na lastnem strežniku. *"Tako občutljivi podatki ne gredo iz varnega in zaščitene okolja,"* je poudaril Robnik Šikonja.

### **Slovenščina brez svojega jezikovnega modela ni suveren jezik**

Če o slovenščini odločajo ameriška podjetja, ta po oceni Centra za jezikovne vire in tehnologije ni suverena. Priprava nacionalnega jezikovnega modela pa bo poleg razvojne neodvisnosti omogočila učinkovit nadzor nad besedili, na katerih bo model naučen, ter ustrezno upoštevanje zakonov o varstvu zasebnosti in osebnih podatkov. Model bo prav tako bolje poznal nacionalne specifikke, kar je ključno za ustrezno in učinkovito komunikacijo.

Največji komercialni modeli, kot sta ChatGPT in Gemini, niso bili učeni na dovolj slovenskih besedilih, zato slovenščino obvladajo slabše. Še šibkejši v slovenščini so prostodostopni modeli, kot so Llama, Mistral, DeepSeek in Gamma, navaja Robnik Šikonja.

Zato nameravajo razviti več generativnih modelov slovenščine (GaMS), ki bodo popolnoma brezplačni in primerni za integracijo v medicini, industriji ter kot novi jezikovni viri za pisni in govorni slovenski jezik. Po oceni centra bo to spodbudilo nadaljnji razvoj in konkurenčnost orodij ter storitev v slovenščini.

**Za učenje modela je potreben superračunalnik, za izvajanje bo zadostoval osebni**

V načrtu imajo modele z eno, devetimi in 27 milijardami parametrov, ki jih trenutno učijo na superračunalniku Euro HPC Leonardo Booster v Bologni. Modeli GaMS bodo sicer manjši in šibkejši od največjih komercialnih modelov, ki po ocenah profesorja obsegajo okoli 1000 milijard parametrov.

A tudi to se utegne kmalu spremeniti, saj je bila Slovenija nedavno uspešna na evropskem razpisu za superračunalnik in tovarno umetne inteligence, ki bo za razliko od obstoječega superračunalnika Vega opremljen z več grafičnimi procesnimi enotami. *"Modeli GaMS bodo na novem superračunalniku dobili svoje naslednike. Mladi GaMS-i pa bodo večji, sposobnejši in bolj prilagojeni posameznim potrebam,"* napoveduje Robnik Šikonja.

Superračunalnik je potreben za učenje modela, ko pa je ta enkrat naučen na veliki količini besedil, zahteva precej manj strojne moči. *"Za osebno rabo in manjše modele bo zadostoval že prenosnik, za večje modele in številne zahteve pa bo potrebnih več zmogljivih grafičnih procesorjev,"* ponazarja. Na centru dodajajo, da bo za jezikovni model z eno milijardo parametrov zadostovala že običajna grafična kartica s približno štirimi gigabajti spomina.

### Javnost pozivajo k oddaji besedil v slovenščini

Različico modela GaMS-1B-Chat, ki je učen na besedilih s približno 10 milijardami besed in ima milijardo parametrov, je mogoče preizkusiti na spletni strani [povejmo.si](https://povejmo.si). V centru poudarjajo, da je model še v prototipni fazi in tisočkrat manjši od najnovejše verzije ChatGPT, zato se v odgovorih tudi pogosteje moti.

Center za jezikovne vire in tehnologije zato poziva založbe, medije in posameznike, naj s prispevanjem besedil aktivno sodelujejo pri gradnji velikega jezikovnega modela. *"Naš cilj je, da ima od velikega jezikovnega modela korist celotna slovenska jezikovna skupnost,"* ob tem tudi bralce Svet24 pozivajo, naj svoja besedila oddajo na spletni strani [zbiranje.povejmo.si](https://zbiranje.povejmo.si).



Profesor računalniških znanosti Marko Robnik Šikonja | Foto: FRI UL