



Jezik in umetna inteligenca

prof. dr. Špela Vintar

Univerza v Ljubljani

spela.vintar@ff.uni-lj.si

- 2016 He was still gazing at the mirror fragment in which, for a split second, he had thought he saw Dumbledore's eye.
- 2018 He was still gazing at the mirror fragment in which, for a split second, he had thought he saw Dumbledore's eye.
- 2023 He was still gazing at the mirror fragment in which, for a split second, he had thought he saw Dumbledore's eye.

Buazizi je bil še vedno ozirati na zrcalni fragmentov, ki pa po drugi, je torej misli Dumbledore svobodne razprave brez seznama govornikov.




Še vedno je gledal v ogledalo fragmenta, v katerem je za delček sekunde, bi bil mislil, da gledajo, Dumbledore oko.

Še vedno je gledal v drobce ogledala, v katerem se mu je za delček sekunde zazdelo, da vidi Dumbledorjevo oko.



Kakšen odnos je imel Harry Potter do svojega bratranca Dudleyja? 



V seriji knjig o Harryju Potterju, ki jih je napisala J.K. Rowling, je bil odnos med Harryjem Potterjem in njegovim bratrancom Dudleyjem Dursleyjem večinoma sovražen. Harry je živel z družino Dursley, ki so ga zanemarjali in zlorabljali, saj so bili njegovi starši umrli, ko je bil še dojenček. Dudley, ki je bil razvajen in razsipen, je pogosto nadlegoval in ustrahoval Harryja.   

Kljub temu pa se njun odnos nekoliko izboljša proti koncu serije. V šesti knjigi, "Harry Potter in Princ mešane krvi," Dudley začne spoznavati, da je njegovo vedenje do Harryja bilo napačno, in se poskuša spremeniti. V zadnji knjigi, "Harry Potter in Relikvije smrti," Dudley celo izrazi nekakšno zaskrbljenost za Harryja, preden se ločijo, ko mora Harry zapustiti domačo hišo.

Tako se odnos med Harryjem in Dudleyjem sčasoma razvije iz sovraštva in zlorabe do neke mere spoštovanja in razumevanja.

Pregled vsebine

- Kako se učijo jezikovni modeli?

- Kako deluje modeliranje jezika z nevronskimi mrežami?

- Kako je strojno prevajanje povezano z umetno inteligenco?

- Kaj je pravzaprav ChatGPT in kako ga uporabljati?

- Kaj pričakovati od razvoja umetne inteligence?

B E H I N D

E V E R Y

C O D E

I S A N

E N I G M A

BENEDICT CUMBERBATCH KEIRA KNIGHTLEY

THE IMITATION GAME

COMING SOON

Alan Turing (1912-1954)

- britanski matematik, logik in kriptograf
- med II. svet. vojno vodil tajno enoto za dešifriranje nemških sporočil; razvil *Ultra*
- velja za očeta sodobnega računalništva in umetne inteligence; *Turingov stroj* pa predstavlja prvi abstraktni model računalniškega algoritma
- "Ali lahko stroji mislijo?" -> Turingov test



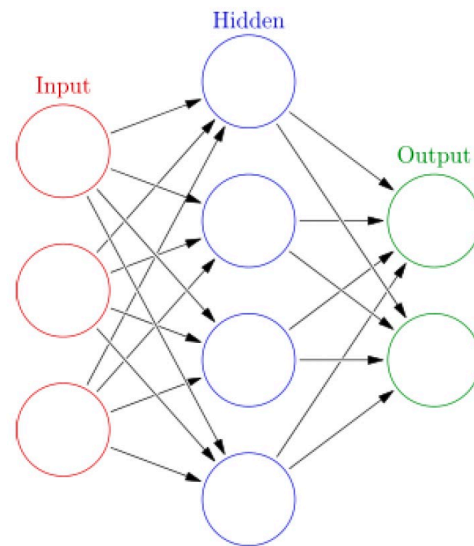
Mejniki zadnjih let

- 2006: Googlov prvi **statistični prevajalnik** na spletu
- 2008: Google med podprte jezike uvrsti tudi **slovenščino**
- 2016: Google pri prevajanju prične uporabljati **nevronske mreže**, in sicer tipa RNN z modelom LSTM (*long short-term memory*)
- 2017: "Attention is all you need" (Vaswani et al. 2017) – nevronska mreža ***transformer***
- 2018: Google zgradi nevronske prevajalnik tudi za slovenščino
- 2018: Na nevronske model prevajanja preidejo vsi večji akterji, tudi prevajalnik EU **eTranslation** (bivši MT@EC)
- 2021: **DeepL** začne podpirati tudi slovenščino
- november 2022: OpenAI objavi **ChatGPT**
- marec 2023: OpenAI objavi **GPT4**
- april 2023: Google objavi svoj jezikovni model **Bard** (dostopen le v ZDA in VB)

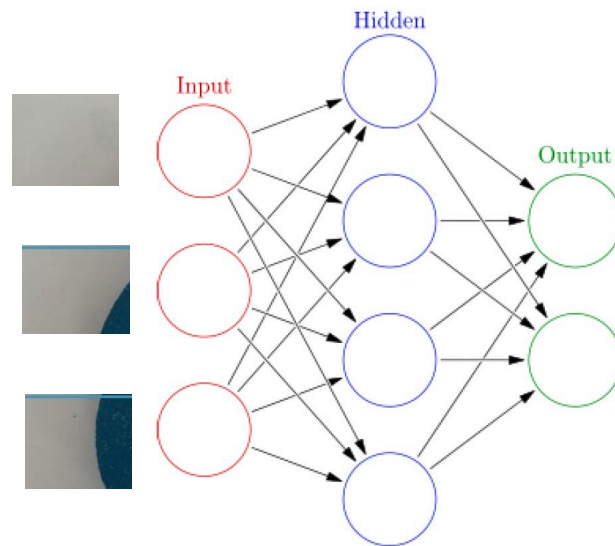
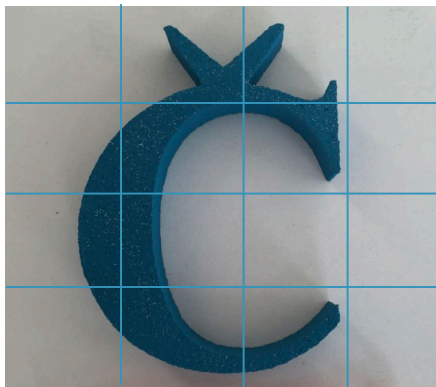
Kako se učejo jezikovni modeli?



Nevronská mreža

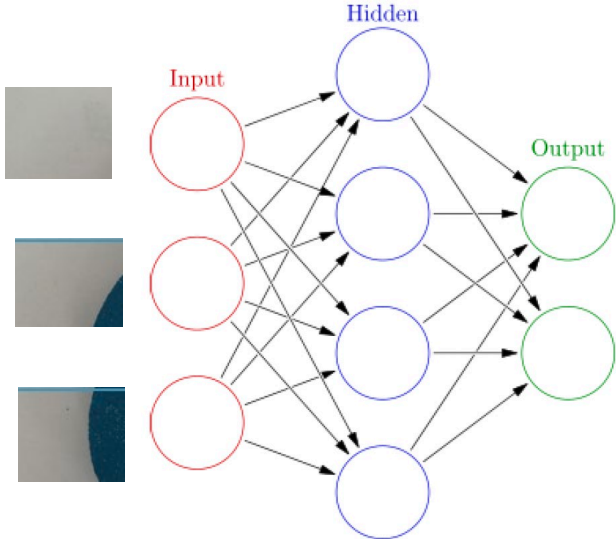
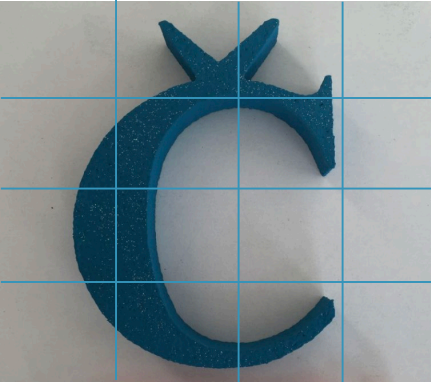


Nevronská mreža



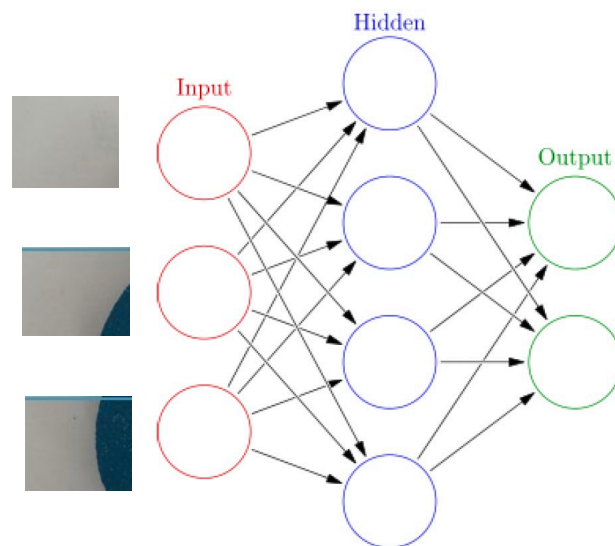
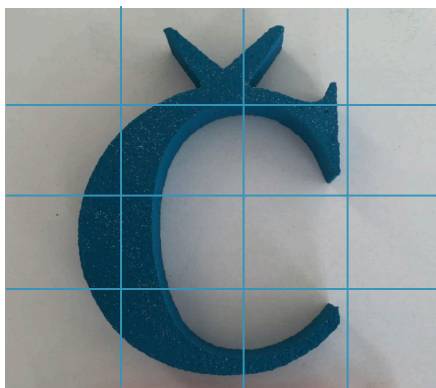
→ b

Nevronska mreža



→ ~~b~~

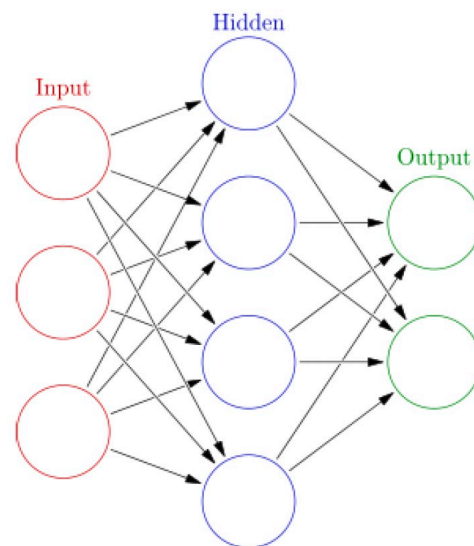
Nevronská mreža



č

Nevronska mreža

Danes je _____ dan.
_____ je deževen dan.
Danes je deževen ____ .



lep
deževen
Igor
Danes
avto
dan

**Kako deluje modeliranje
pomena z nevronskimi
mrežami?**

Distribucijska semantika

- "You shall know the word by the company it keeps."

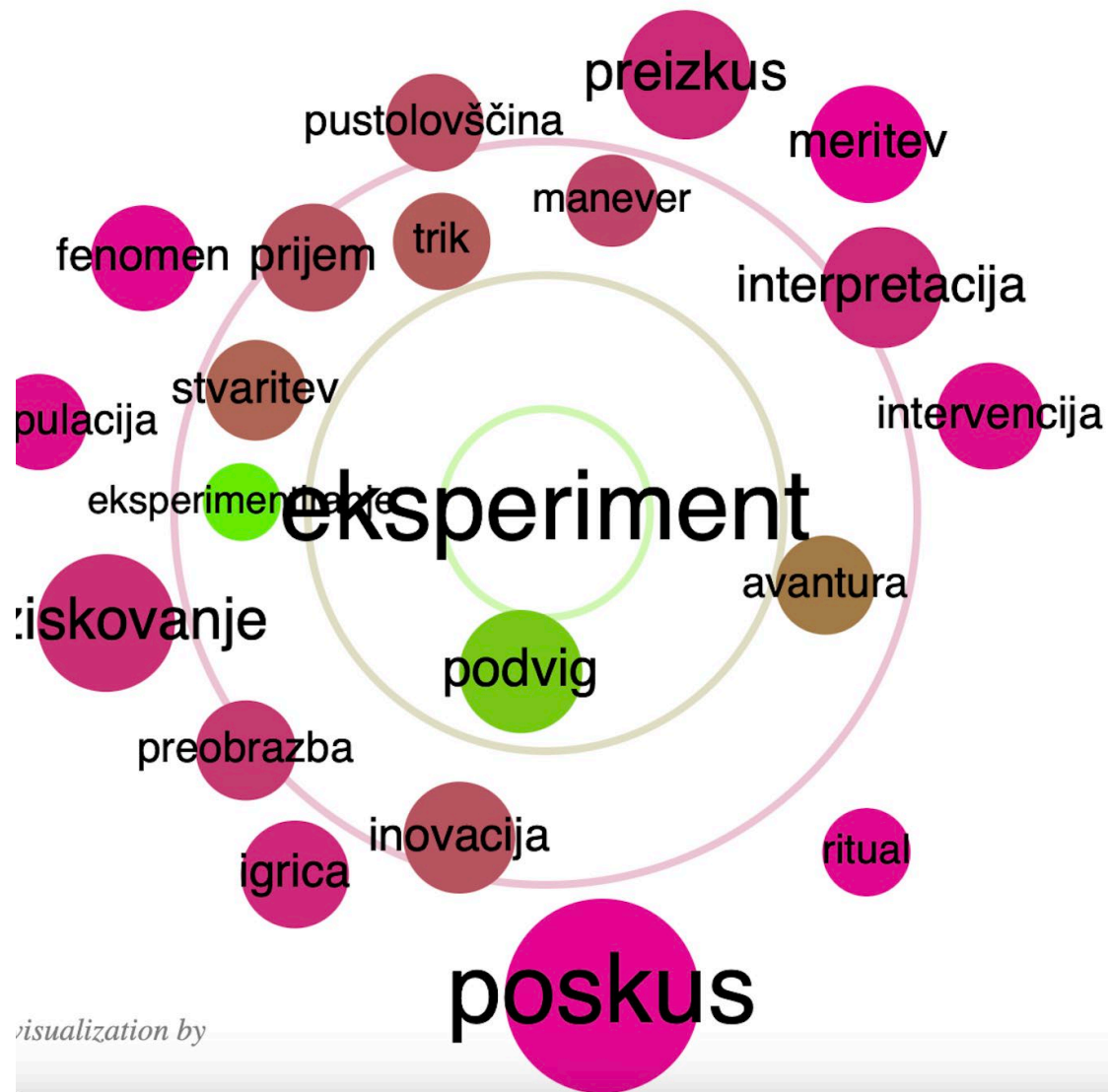
Firth 1957

in diskusija ¶ Namen naloge je bil na podlagi **eksperimentov** , izvedenih v laboratorijskem merilu, koncentraciji) tako, da smo pri posameznih **eksperimentih** uspeli pridobiti le po tri meritve, ki pa so kinetičnih parametrov. Na osnovi izvedenih **eksperimentov** lahko z veliko verjetnostjo trdimo, da kalijev , najnižjo smo dosegli pri pH = 11. Nadaljne **eksperimente** smo zato izvajali pri tej pH vrednosti. , je bila temperatura reakcije. Vse nadaljnje **eksperimente** smo izvajali pri predhodno določenem pH = 11. amoksicilina in klavulanske kisline. Tekom **eksperimentov** smo ugotovili, da bomo lahko analize izvajali ¶ Na podlagi rezultatov, dobljenih z številnimi **eksperimenti** , lahko potrdimo, da najnižjo koncentracijo , s katerim lahko amoksicilin reagira. Z **eksperimenti** smo to tudi potrdili. ¶ Na podlagi rezultatov smo

Distribucijska semantika

in diskusija ¶ Namen naloge je bil na podlagi **eksperimentov** , izvedenih v laboratorijskem merilu, koncentraciji) tako, da smo pri posameznih **eksperimentih** uspeli pridobiti le po tri meritve, ki pa so kinetičnih parametrov. Na osnovi izvedenih **eksperimentov** lahko z veliko verjetnostjo trdimo, da kalijev , najnižjo smo dosegli pri pH = 11. Nadaljne **eksperimente** smo zato izvajali pri tej pH vrednosti. , je bila temperatura reakcije. Vse nadaljnje **eksperimente** smo izvajali pri predhodno določenem pH = 11. amoksicilina in klavulanske kisline. Tekom **eksperimentov** smo ugotovili, da bomo lahko analize izvajali ¶ Na podlagi rezultatov, dobljenih z številnimi **eksperimenti** , lahko potrdimo, da najnižjo koncentracijo , s katerim lahko amoksicilin reagira. Z **eksperimenti** smo to tudi potrdili. ¶ Na podlagi rezultatov smo

	namen	naloge	je	bil	podlagi	izvedenih	merilu
eksperiment	1	1	2	1	3	2	1



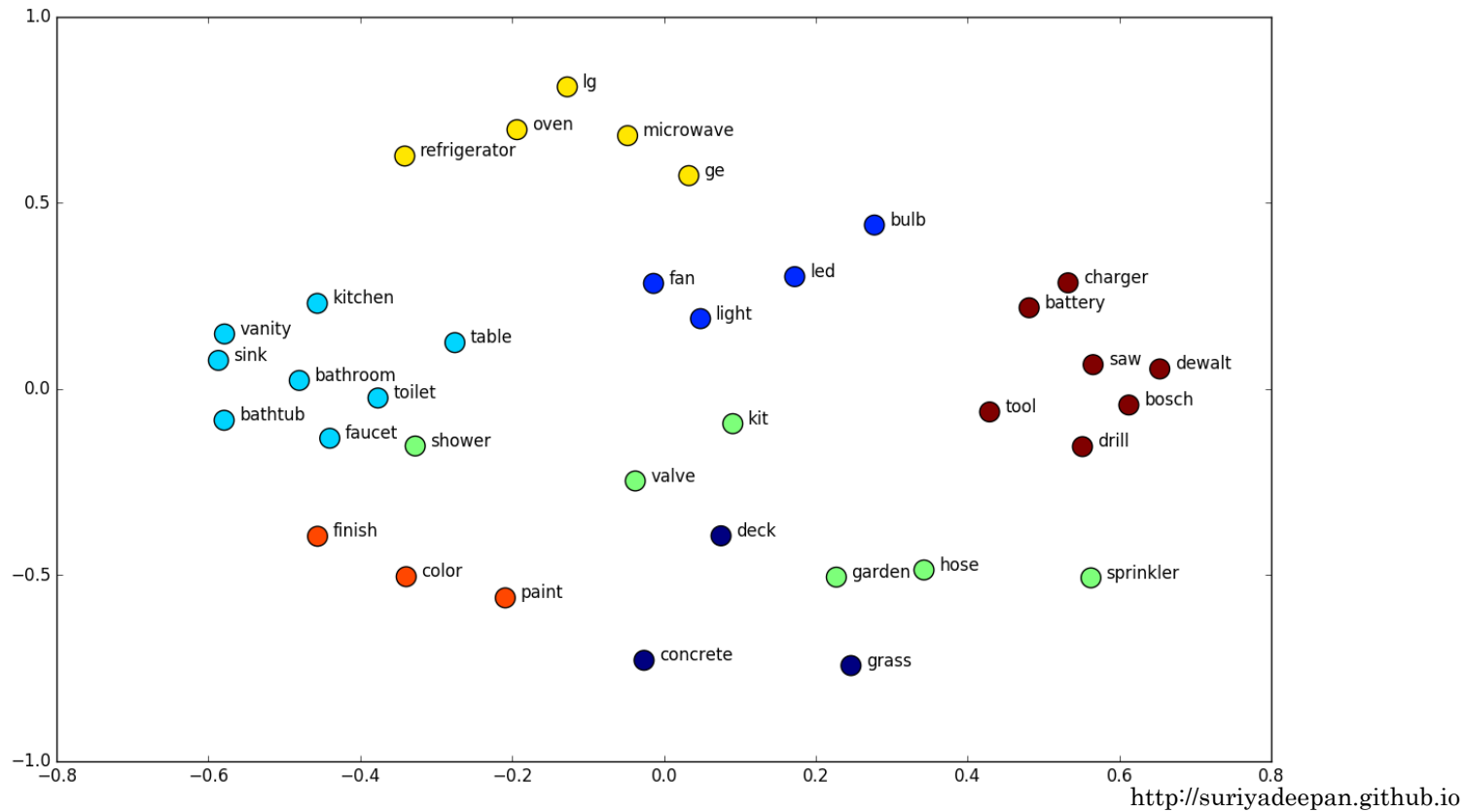
visualization by

Besedni oblak

Vektorska vložitev

eksperiment 0.157417 0.013155 -0.168669 0.258004 -0.261520 -0.004266 0.064963 -0.095019 0.110000 0.139827 -0.009515
0.102139 -0.014292 0.064913 0.116374 -0.248112 -0.179980 0.143360 -0.058822 -0.081025 -0.204884 0.024653 -0.008157 -0.208811
-0.351810 -0.020040 -0.191807 -0.004969 0.148161 -0.124785 0.073630 -0.063658 -0.325918 0.170596 0.123046 0.106464 -0.141023
0.037222 0.095813 0.349635 0.166654 -0.160324 -0.036104 0.016911 0.002022 0.026284 -0.129991 -0.061656 0.075007 -0.200100 -
0.026334 -0.169233 0.033682 0.031311 0.034157 -0.056466 -0.010009 0.373934 -0.461444 0.137428 -0.080984 -0.214958 0.133438 -
0.206060 -0.044642 -0.020855 0.001350 -0.032700 0.054392 -0.138733 0.271986 0.074465 -0.383632 -0.060489 0.170901 -0.007525
-0.213951 -0.215574 -0.237169 0.047382 0.059949 -0.167393 -0.086040 0.037600 -0.064871 0.099520 -0.137147 0.170327 -0.044044
0.238244 0.082285 -0.030852 0.061542 -0.099835 -0.154609 -0.162004 0.061578 -0.172150 -0.163639 -0.012122 -0.085171 0.197971
0.483274 0.020190 0.203729 0.267945 -0.159703 -0.352687 0.009871 0.032857 0.076368 0.090400 0.143115 0.161378 -0.203485 -
0.020512 -0.139726 0.174444 -0.035236 0.069389 0.157347 -0.254437 0.016946 -0.190948 -0.186738 0.060985 0.376371 -0.084888 -
0.330832 -0.032036 0.072837 -0.381461 0.088491 -0.315386 0.119776 -0.174684 0.337164 0.031623 0.036908 0.297419 -0.208412
0.316575 0.036308 0.119306 -0.025103 0.110816 -0.072111 -0.197073 0.154872 0.285448 0.063191 -0.018339 -0.117289 -0.089172
0.033924 -0.297626 -0.094350 -0.106417 -0.283593 -0.126537 0.131320 -0.087875 -0.044319 -0.143214 -0.086091 0.147566
0.043141 -0.121772 -0.123281 -0.044752 0.036134 -0.273063 -0.046211 0.014122 -0.001258 0.216371 0.027509 -0.117638 0.249159
0.137707 0.087296 0.059431 0.029131 -0.079217 0.007073 0.006122 -0.015539 0.112570 0.250844 -0.012353 -0.162030 -0.107101
0.072732 -0.159938 0.064535 0.069997 0.155462 -0.015620 0.390249 0.096999

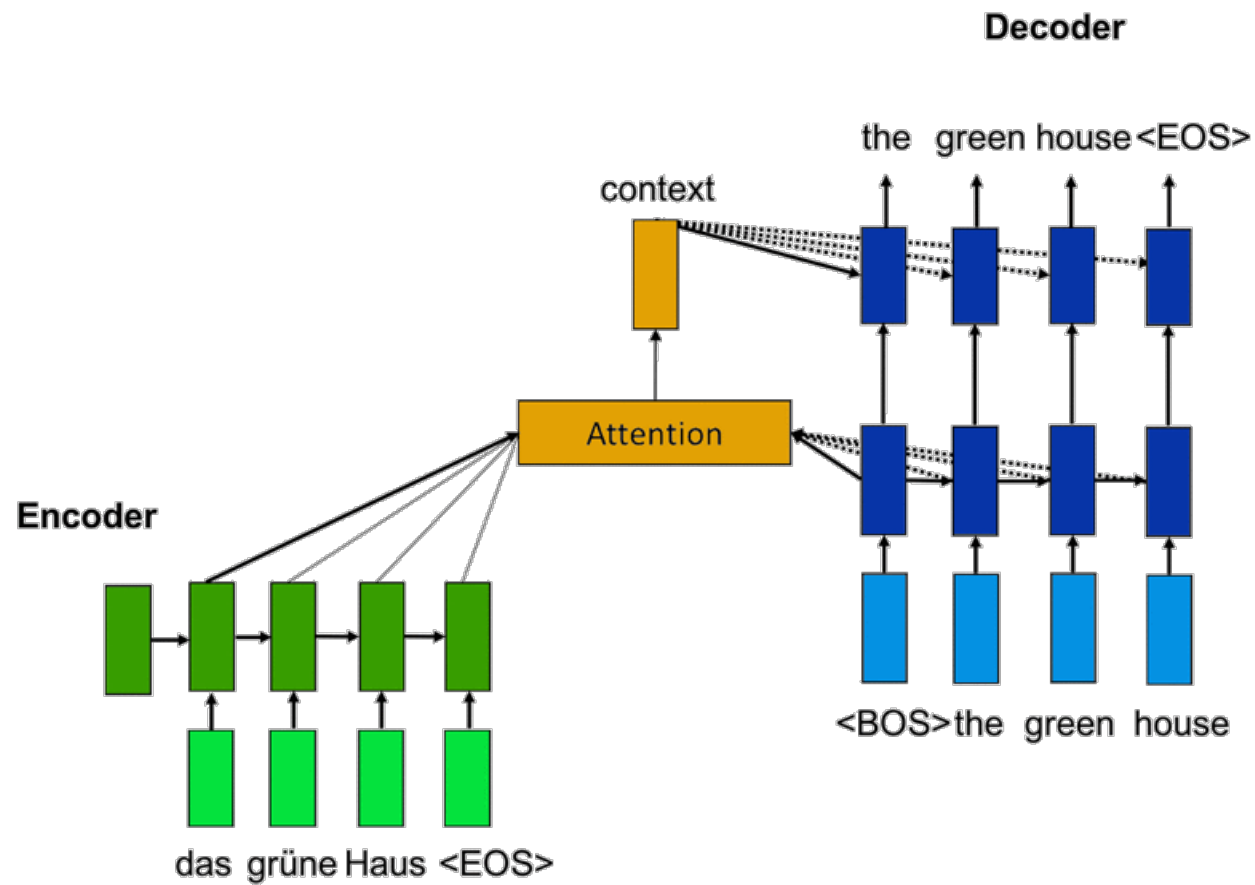
Vektorske vložitve v dvodimenzionalnem prostoru



**Kako je strojno
prevajanje
povezano z
umetno
inteligenco?**



Nevronsko strojno prevajanje



Nekaj podatkov

- Na konferenci WMT2018 sta dva prevajalnika prvič prekosila človeka. Na WMT2019 je bilo takšnih sistemov že prek 10.
- Z uporabo prevajalnikov se produktivnost človeških prevajalcev v povprečju poveča poveča za 400 %.
- Raziskava ELIS:

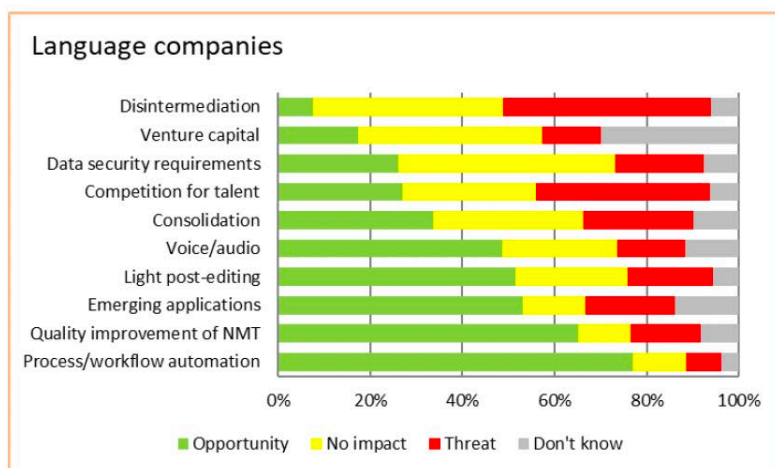


Figure 12 - Trend opinions - language companies

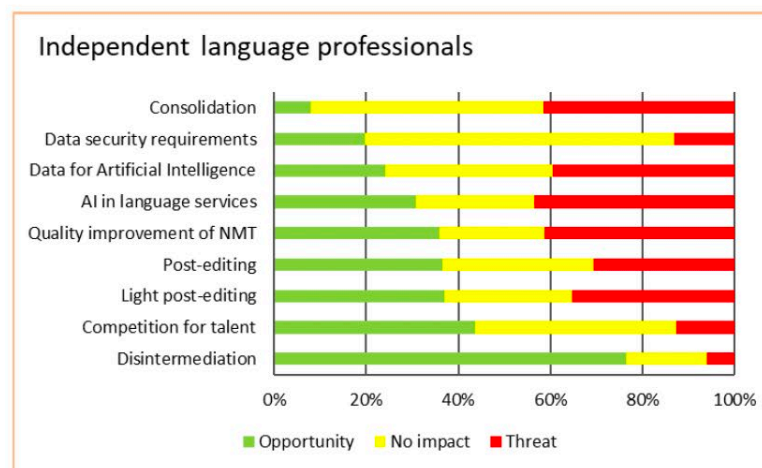


Figure 13 - Trend opinions - independent professionals

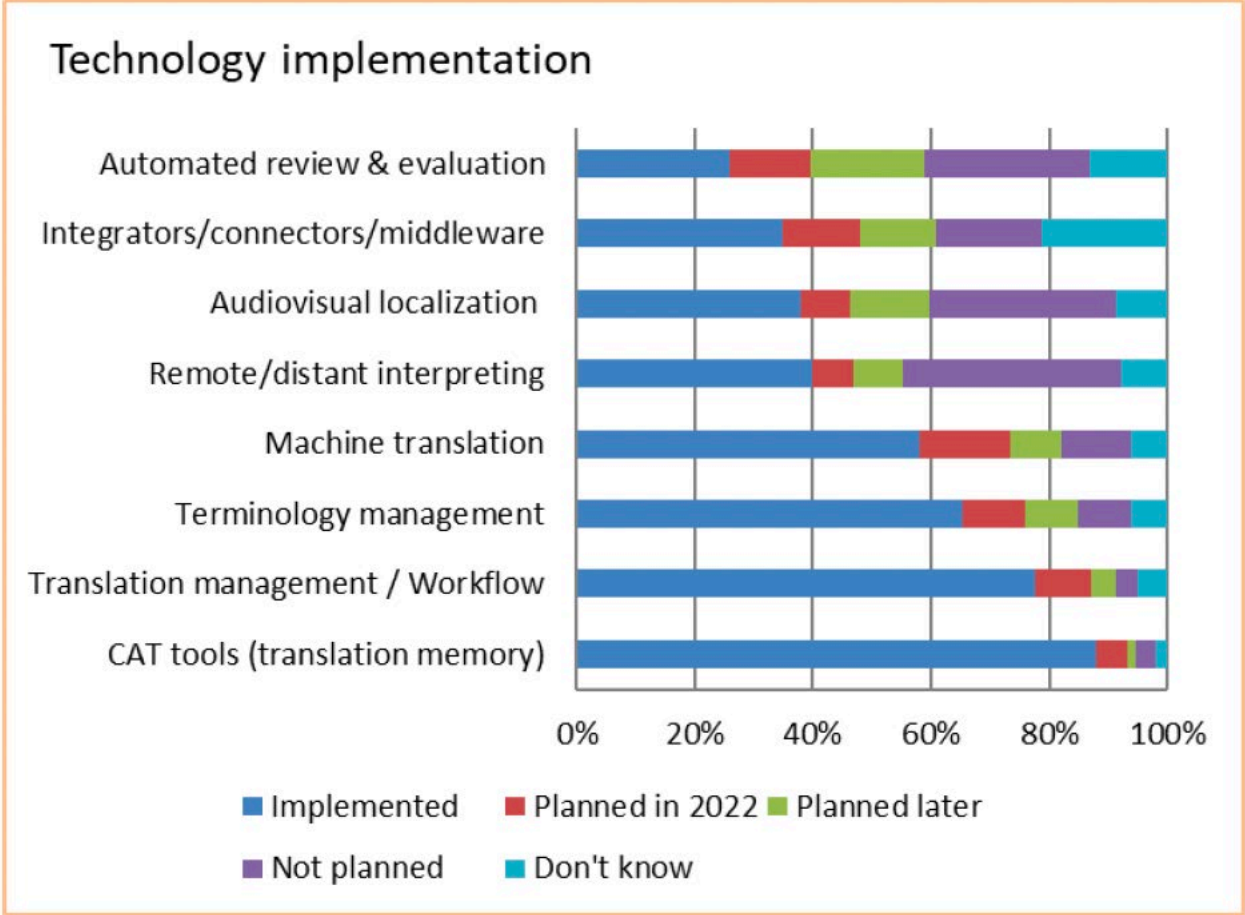


Figure 43 - Technology implementation in language companies



Izzivi, ki še ostajajo

- prevajanje književnih in slogovno **zahtevnih** besedil
- doslednost pri prevajanju **izrazja**, imen itd.
- tehnično **nedovršena prevajalska orodja**
- lokalizacija, **transkreacija**
- strojno **tolmačenje**
- ...



Zdaj bova isto počela v slovenščini. Dolžina metafore zaenkrat ni omejena.
Na steni je sedel ogromen pajek, ki me je spominjal na



gotško kapelo, prežeto s senčnimi vzorci in čudovito grozo, v kateri je vsaka noga predstavljala vejo v tem, nenavadnem labirintu.



Da sem se zaljubila v tega tipa je bilo v tistem trenutku podobno



kot da bi se prepustila divji reki, ki je brezusmilno nesla moje srce na neznano, nepredvidljivo pot.



brezusmilno?

ChatGPT & co.

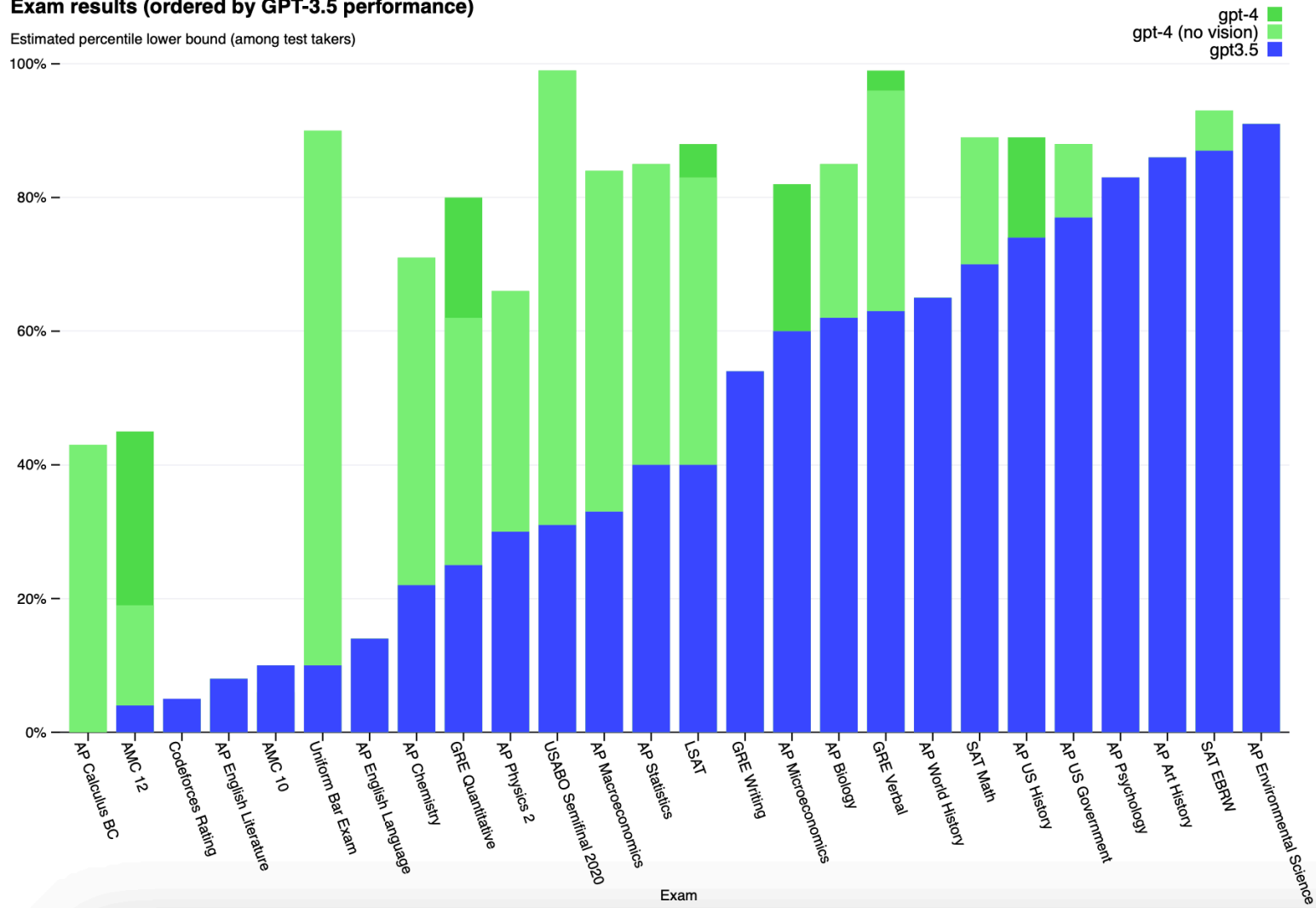
Veliki jezikovni modeli

- napovedni model, ki predvideva zaporedja besed ali besednih zvez v jeziku, naučen na (pretežno) jezikovnih podatkih

A	B	C	D	E	F	G	H	I
Dataset	Total size (GB)	Wiki	Books/Journals	Discussion	Code	Web (CC/C4)	Announced	Model examples
RedPajama (Together AI)	4033	80	180	67	197	3510	Apr/2023	MPT
Stability New Pile (Stability AI)	5000	-	-	-	-	-	Apr/2023	StableLM
GPT-4 (OpenAI)	40000	-	-	-	-	-	Mar/2023	GPT-4
LLaMA (Meta AI)	4749	83	177	78	328	4083	Feb/2023	LLaMA, Alpaca
MassiveText EN (DeepMind)	10550	12.5	2264.4		3100	5173.1	Dec/2021	Chinchilla, Gopher
Infiniset (Google)	12616	1569	1632	6277	1569	1569	May/2021	LaMDA
The Pile v1 (EleutherAI)	825	6	362	166.71	95	227	Dec/2020	GPT-Neo, GPT-J
GPT-3 (OpenAI)	753	11.4	122			620	May/2020	GPT-3
RoBERTa (Meta AI)	161	11.4	4.6			145	Jul/2019	RoBERTa, Megatron-11B
GPT-2 (OpenAI)	40					40	Feb/2019	GPT-2
GPT-1	4.6		4.6				Jun/2018	GPT-1

Exam results (ordered by GPT-3.5 performance)

Estimated percentile lower bound (among test takers)



Ključni elementi

- količina učnih podatkov za *predučenje*
- število parametrov
- optimiziranje modela z dajanjem navodil
- optimiziranje modela s pogovori (*reinforcement learning from human feedback, RLHF*)
- vrsta optimizacij v zvezi z varnostjo uporabe

ChatGPT in jezik

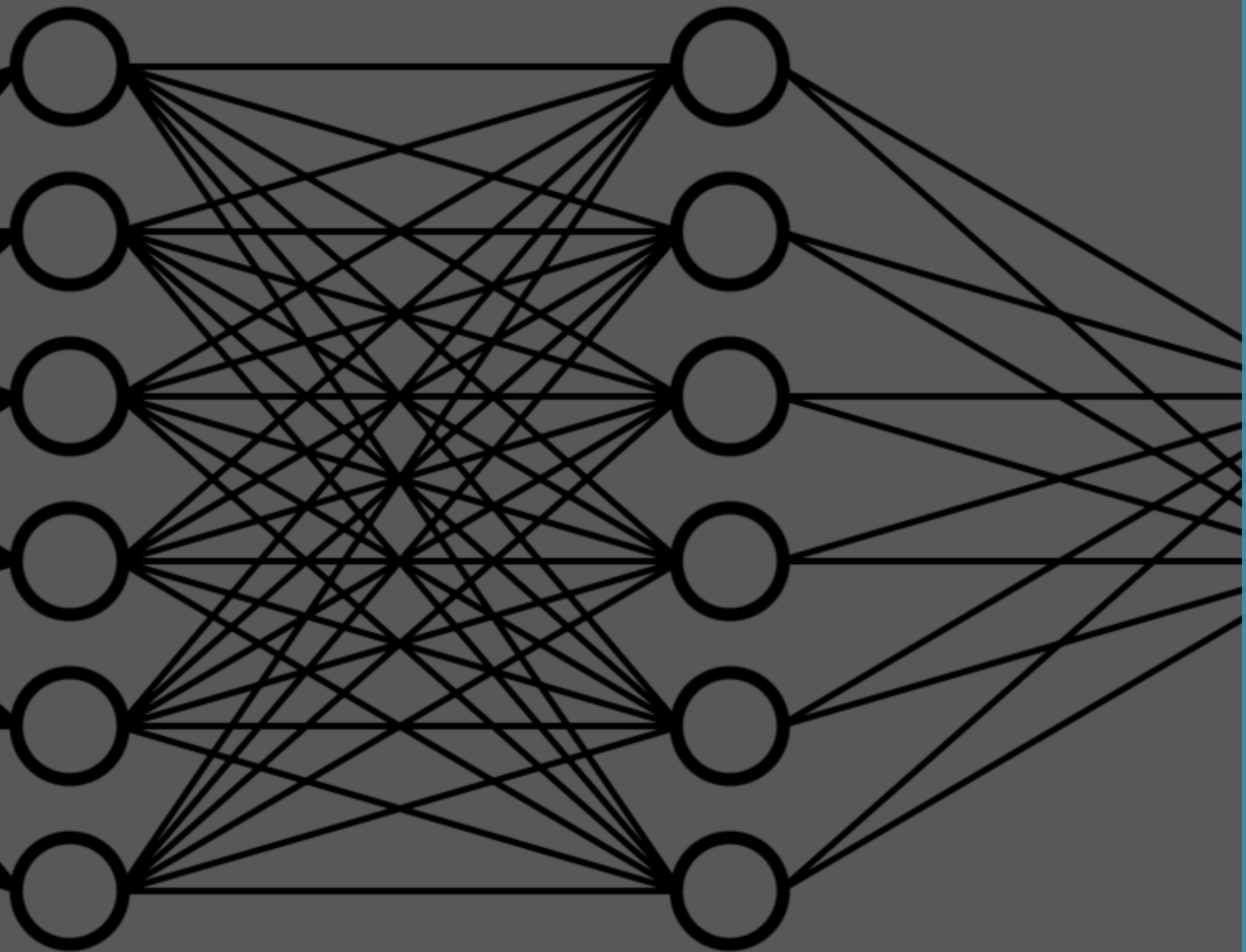
- prevajanje: <https://www.makeuseof.com/how-to-translate-with-chatpt/>
- učenje jezikov: <https://www.scribbr.com/ai-tools/chatgpt-language-learning/>
- preverjanje slovnice: <https://silentpcreview.com/ai/can-chatgpt-check-grammar/>
- ...

Drugi načini uporabe

- medicina,
- pravo,
- računalništvo,
- izobraževanje,
- ...

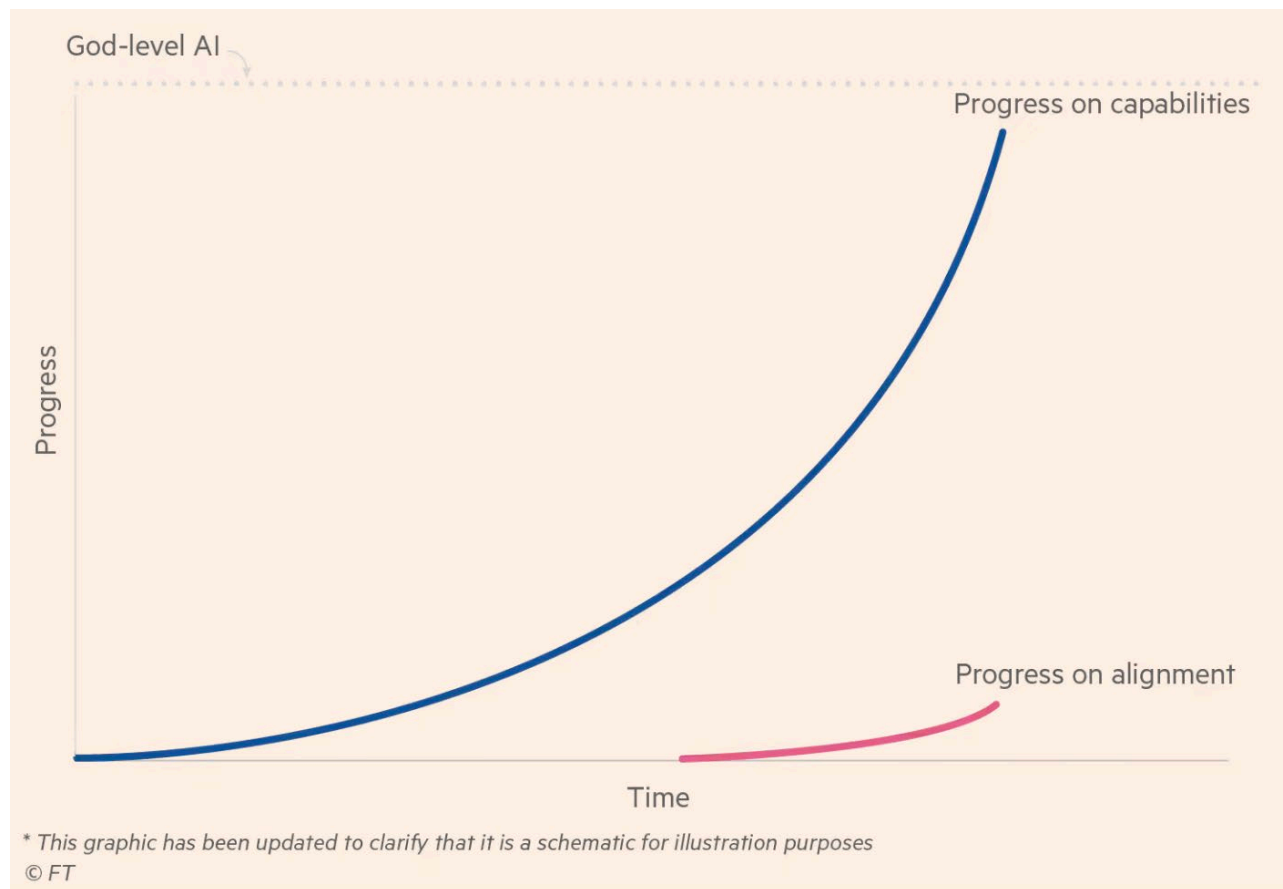
Izzivi

- varstvo podatkov
- etika uporabe
- pristranskost
- enakopravnost
- transparentnost
- zanesljivost



O prihodnosti

- We Must Slow Down the Race...(Financial Times, 13. 4. 2023)
<https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2>
- EU AI Act (2021 ->)
[https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en)



Viri

- ANTAKI, Fares, TOUMA, Samir, MILAD, Daniel, EL-KHOURY, Jonathan, DUVAL, Renaud, 2023: Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *medRxiv*, 2023-01.
- BARRAULT, Loïc, BOJAR, Ondřej, COSTA-JUSSA, Marta R., FEDERMANN, Christian, FISHEL, Mark, GRAHAM, Yvette, HADDOW, Barry et al., 2019: Findings of the 2019 conference on machine translation (WMT19). Association for Computational Linguistics.
- BOJAR, Ondřej, FEDERMANN, Christian, FISHEL, Mark, GRAHAM, Yvette, HADDOW, Barry, HUCK, Matthias, KOEHN, Philipp, MONZ, Christof, 2018: Findings of the 2018 Conference on Machine Translation (WMT18). Proceedings of the Third Conference on Machine Translation. Association for Computational Linguistics. 272–307.
- do CAMPO BAYÓN, María, SÁNCHEZ-GIJÓN, Pilar, 2022: Evaluating NMT: superior, inferior, or equivalent to texts originally written by humans. *New Trends in Translation and Technology NeTTT 2022*, 4-6 July 2022, Rhodes Island, Greece. 27-36.
- CARO QUINTANA, Rocío, CASTILHO, Sheila, 2022: A review of the Integration of Machine Translation in CAT tools. In *New Trends in Translation and Technology 2022 NeTTT 2022*, 4-6 July 2022, Rhodes Island, Greece. 214-221.
- European Language Industry Survey 2022: Trends, expectations and concerns of the European language industry. ELIA, https://fit-europe-rc.org/wp-content/uploads/2022/03/ELIS-2022_survey_results_final_report.pdf?x85225.
- FRIEDER, Simon, PINCHETTI, Luca, GRIFFITHS, Ryan-Rhys, SALVATORI, Tomasso, LUKASIEWICZ, Thomas, PETERSEN, Philip Christian, CHEVALIER, Alexis, BERNER, Julius, 2023: Mathematical capabilities of ChatGPT. arXiv preprint: 2301.13867.
- JEBLICK, Katharina, SCHACHTNER, Balthasar, DEXL, Jakob, MITTERMEIER, Andreas, STÜBER, Anna Theresa, TOPALIS, Johann, WEBER, Thomas, WESP, Philipp, SABEL, Bastian, RICKE, Jens, in dr., 2022: Chatgpt makes medicine easy to swallow: An exploratory case study on simplified radiology reports. arXiv:2212.14882. <https://doi.org/10.48550/arXiv.2212.14882>
- JIAO, Wenxiang, WANG, Wenxuan, HUANG, Jen-tse, WANG, Xing, TU, Zhaopeng, 2023: Is ChatGPT a good translator? A preliminary study. arXiv: 2301.08745. <https://doi.org/10.48550/arXiv.2301.08745>
- KOCMI, Tom, BAWDEN, Rachel, BOJAR, Ondřej, DVORKOVICH, Anton, FEDERMANN, Christian, FISHEL, Mark, GOWDA, Thamme, et al., 2022: Findings of the 2022 conference on machine translation (WMT22). Proceedings of the Seventh Conference on Machine Translation (WMT). 1-45.
- KOUBAA, Anis, BOULILA, Wadii, GHOUTI, Lahouari, ALZAHAM, Ayyub, LATIF, Shahid, 2023: Exploring ChatGPT Capabilities and Limitations: A Critical Review of the NLP Game Changer. Preprints.org, 2023030438. <https://doi.org/10.20944/preprints202303.0438.v1>
- MURGOLO, Elena, 2022: Productivity Evaluation in MT Post-Editing and Fuzzy Matches Ed-iting. Setting the Threshold. *New Trends in Translation and Technology NeTTT 2022*. 4-6 July 2022, Rhodes Island, Greece. 1-8.
- ORTEGA-MARTÍN, Miguel, GARCÍA-SIERRA, Oscar, ARDOIZ, Alfonso, ÁLVAREZ, Jorge, ARMENTEROS, Juan Carlos, ALONSO, Armenteros, 2023: Linguistic ambiguity analysis in ChatGPT. arXiv:2302.06426. <https://doi.org/10.48550/arXiv.2302.06426>
- SUSNJAK, Teo, 2022: ChatGPT: The end of online exam integrity? arXiv: 2212.09292. <https://doi.org/10.48550/arXiv.2212.09292>