



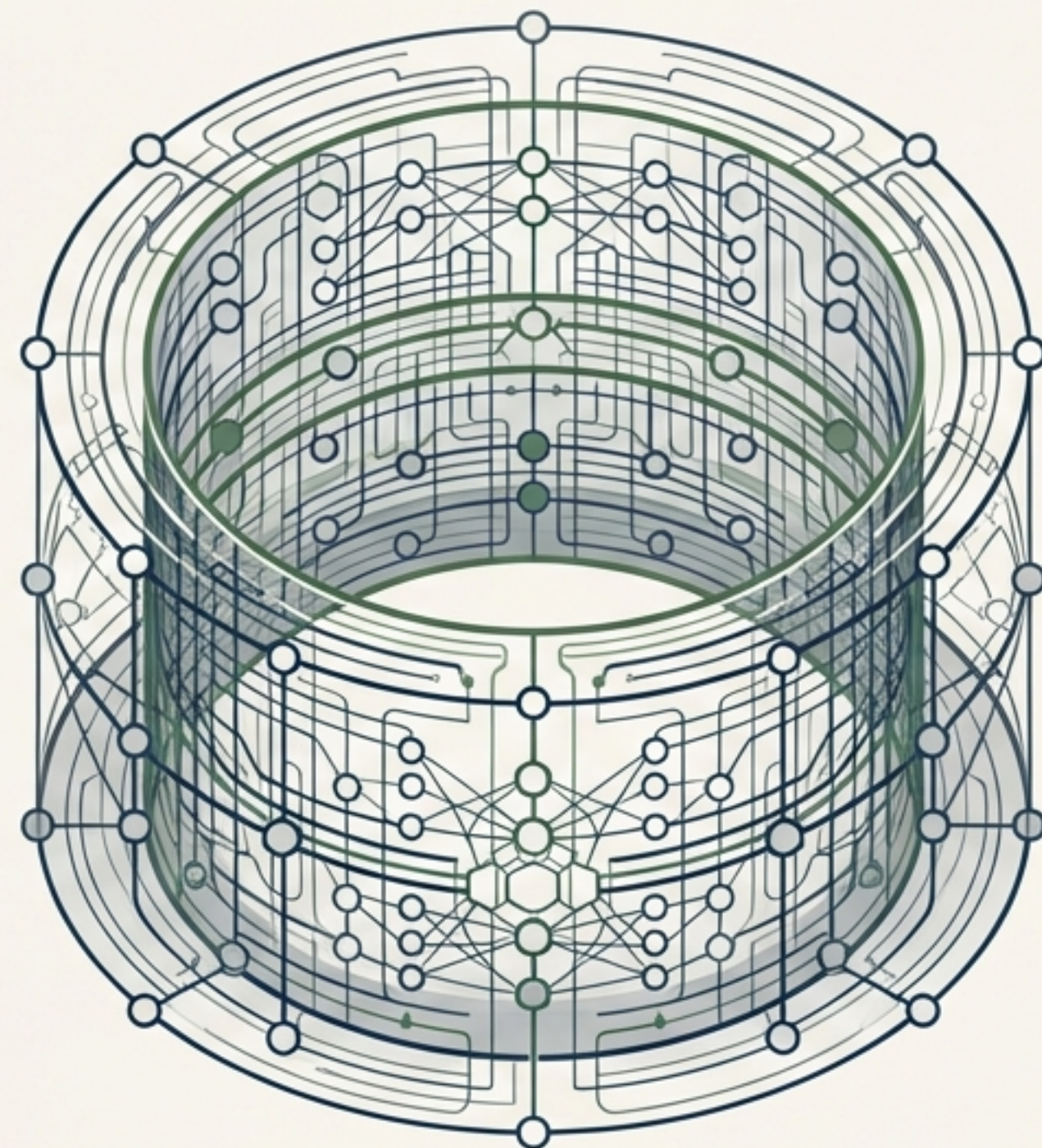
Načrt za dobo po LLM

Razvoj v smeri modelov UI, obogatenih z znanjem,
sodelovalnih in so-razvijajočih se

Na podlagi raziskovalnega članka: Wu et al. (2025), *Engineering*.

Revolucija velikih jezikovnih modelov (LLM)

Veliki jezikovni modeli so korenito pospešili napredek umetne inteligence in dosegli izjemne zmožnosti pri nalogah, kot so razumevanje, **generiranje** in **sklepanje** v več modalnostih.



Vendar pa imajo sedanji LLM–ji neločljive omejitve



Znanje in zanesljivost

- Zastarele ali neaktualne informacije.
- Halucinacije in dejstvene netočnosti.
- Nezmožnost sklepanja nad strukturiranimi podatki.



Učinkovitost in specializacija

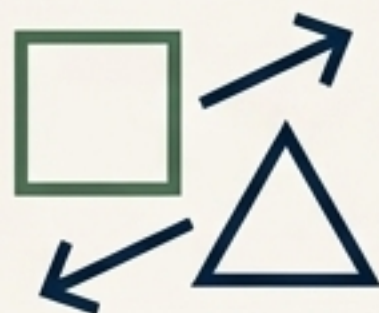
- Ogromne zahteve po računskih virih.
- Nepreglednost in pomanjkanje razločljivosti.
- Nižja natančnost pri specifičnih, domenskih nalogah.



Prilagodljivost in učenje

- Usposobljeni na statičnih naborih podatkov.
- Težave z nenehnim učenjem in prilagajanjem novim informacijam brez obsežnega ponovnega učenja.

Temeljni izzivi za naslednjo generacijo umetne inteligence



Heterogenost nalog

Modeli so razviti za različne naloge z nasprotujočimi si cilji optimizacije.



Heterogenost modelov

Drastične razlike v arhitekturi in notranjih reprezentacijah ovirajo sinergijo.



Heterogenost podatkov

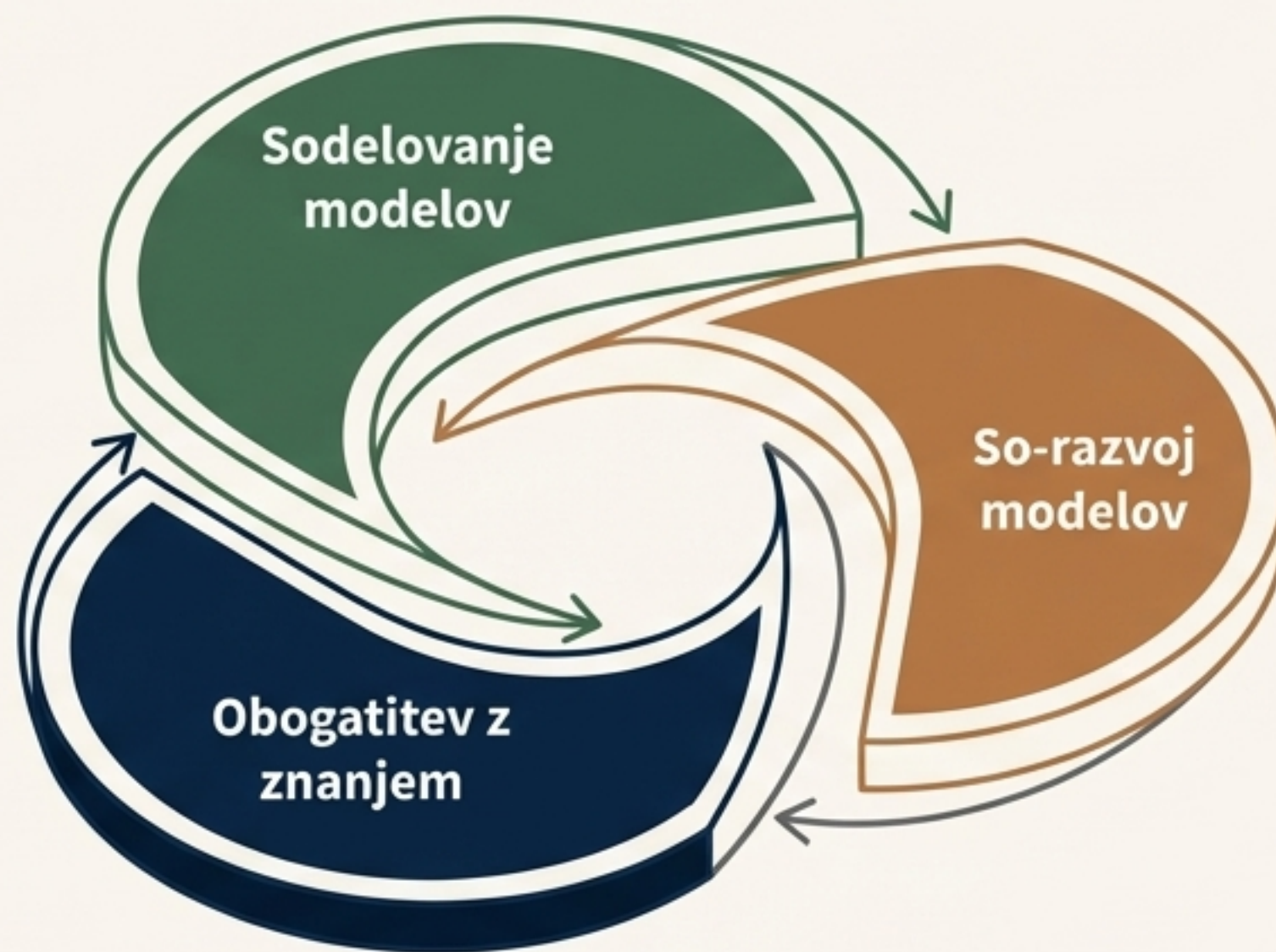
Ne-IID podatki (neodvisni in identično porazdeljeni) iz različnih virov zmanjšujejo zmogljivost posploševanja.



Varnost in zasebnost

Tveganja, kot so rekonstrukcija podatkov iz gradientov, inferenčni napadi in napadi z zastrupljenjem, v porazdeljenih sistemih.

Naš načrt: Trije stebri evolucije umetne inteligence



Skupaj ti stebri tvorijo medsebojno povezan okvir za izboljšanje zmoglosti UI, ki presegajo zmoglosti posameznih modelov.

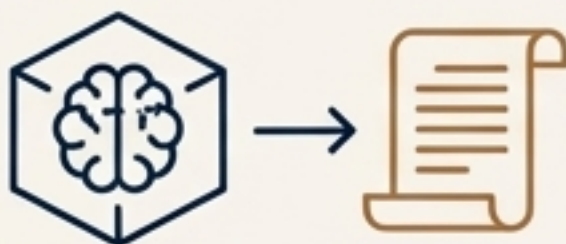
Steber 1: Obogatitev z znanjem

Cilj: Premagati zastarele informacije in halucinacije z integracijo zunanjih, eksplicitnih virov znanja za izboljšanje dejstvene točnosti in razložljivosti.

Ključne metode



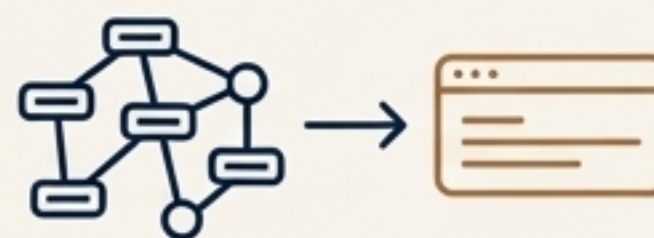
Vključevanje v cilje usposabljanja: Oblikovanje funkcij izgube, ki upoštevajo znanje (npr. višja verjetnost maskiranja za pomembne entitete).



Prilagajanje na podlagi navodil (Instruction Tuning): Fino uravnavanje modelov za razumevanje zunanjega znanja, pretvorjenega v naravni jezik.



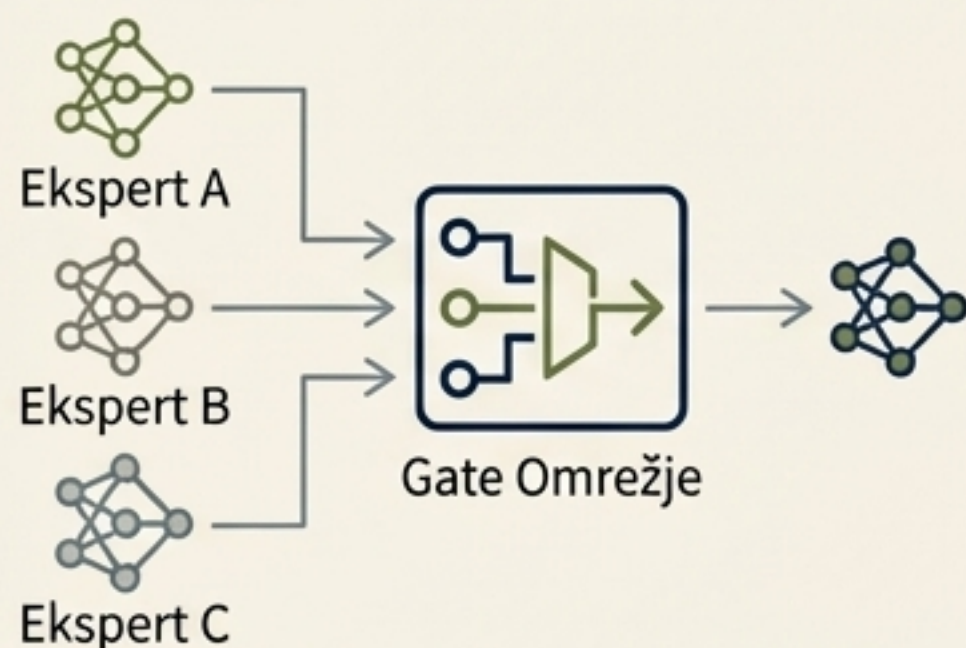
Z iskanjem obogatena inferenca (Retrieval-Augmented Inference): Dinamično iskanje in vključevanje relevantnih informacij med generiranjem odgovora (npr. RAG).



'Knowledge Prompting': Pretvorba strukturiranega znanja (npr. iz grafov znanja) v tekstovne pozive brez ponovnega učenja modela.

Steber 2: Sodelovanje modelov

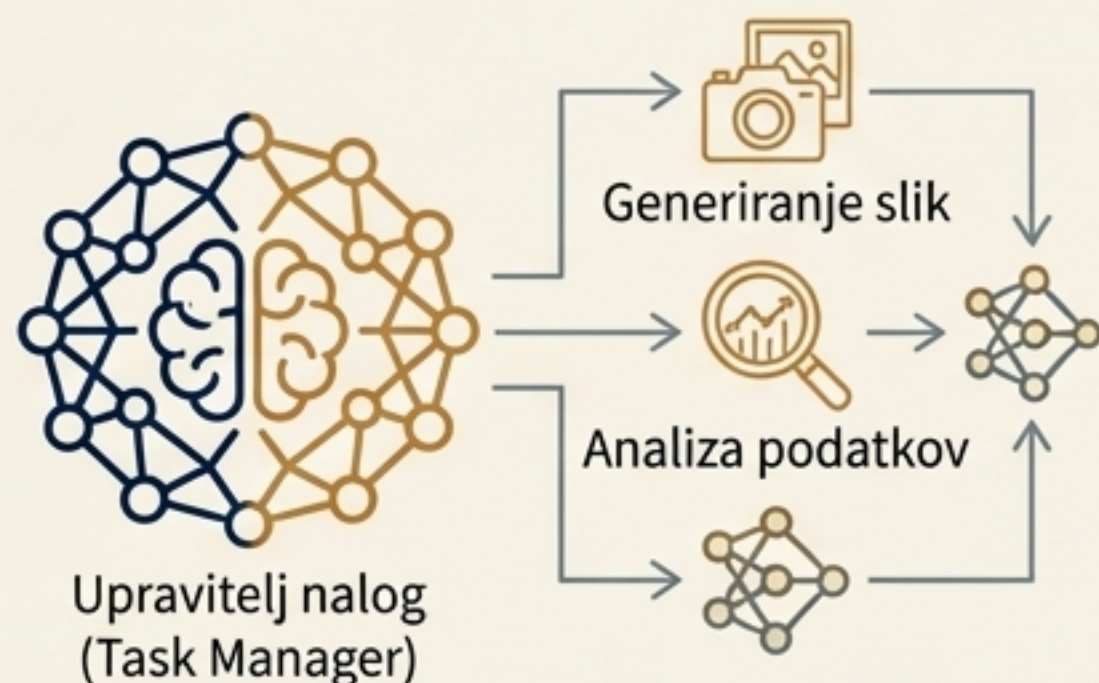
Cilj: Združiti splošne zmožnosti velikih modelov z učinkovitostjo, specializacijo in zanesljivostjo manjših ali funkcionalno različnih modelov.



Zlivanje modelov (Model Merging)

Opis: Kombiniranje napovedi, struktur ali parametrov več modelov za izboljšanje zmogljivosti.

Primer: Arhitektura 'Mešanica strokovnjakov' (Mixture of Experts - MoE), kjer 'gate' omrežje dinamično aktivira specializirane 'ekspertne' modele.



Funkcionalno sodelovanje (Functional Collaboration)

Opis: Inteligentni agentni sistem, kjer en model deluje kot koordinator, drugi pa kot specializirana orodja.

Primer: LLM kot 'upravitelj nalog' (task manager), ki kliče različne specializirane modele (npr. za generiranje slik, analizo podatkov) za reševanje kompleksnih nalog.

Steber 3: So-razvoj modelov

Cilj: Omogočiti dinamičen proces, v katerem se več modelov razvija skupaj, vpliva drug na drugega in si deli znanje, da se prilagodijo spreminjajočim se okoljem in nalogam.

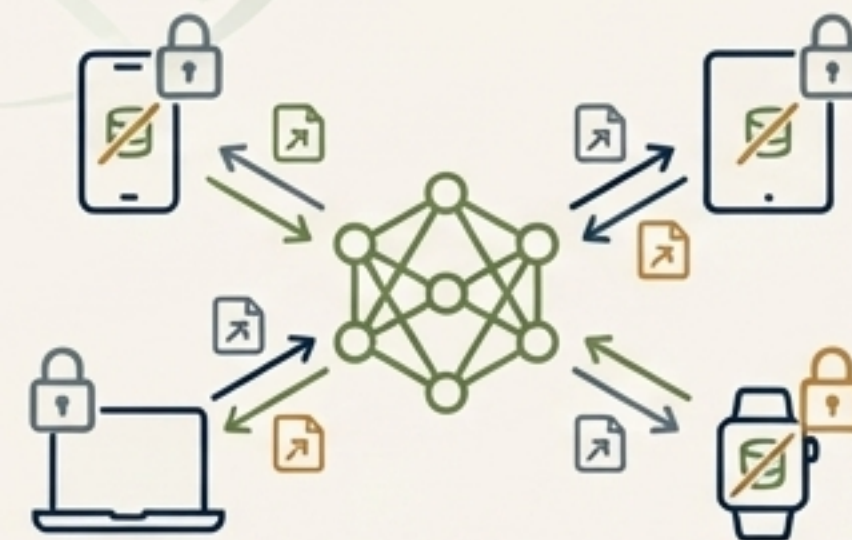
Strategije za premagovanje heterogenosti



Pri modelni heterogenosti:
Dvojna destilacija znanja (Dual KD),
kjer si modeli medsebojno
izmenjujejo vloge učitelja in učenca.

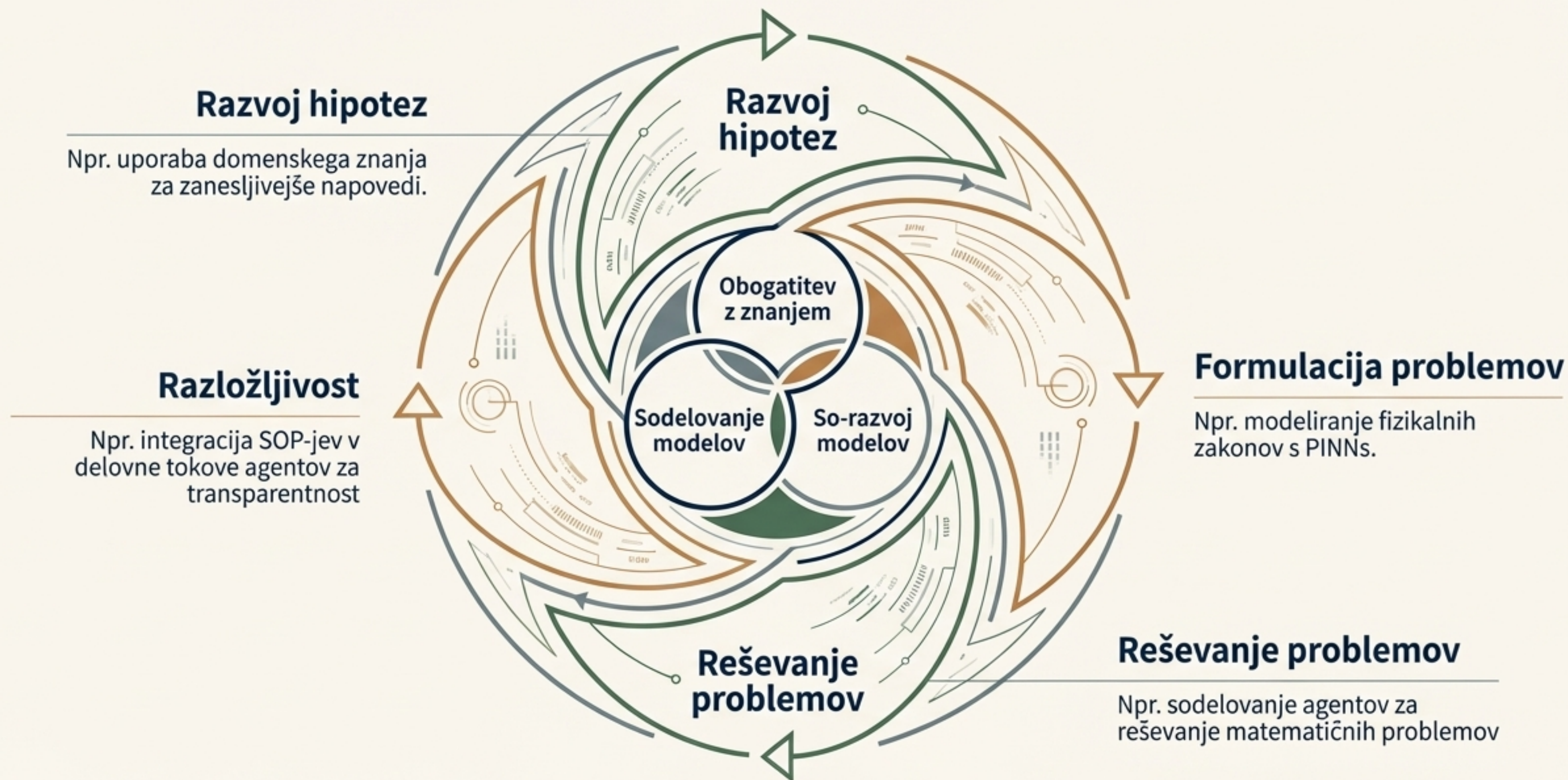


Pri nalogni heterogenosti:
Adversarialno učenje, kjer
tekmovanje med modeli
izboljša robustnost.



Pri podatkovni heterogenosti:
Pri podatkovni heterogenosti:
Federativno učenje, ki omogoča
sodelovanje modelov brez deljenja
surovih, zasebnih podatkov.

Sinergija v praksi: UI za znanost, inženirstvo in družbo

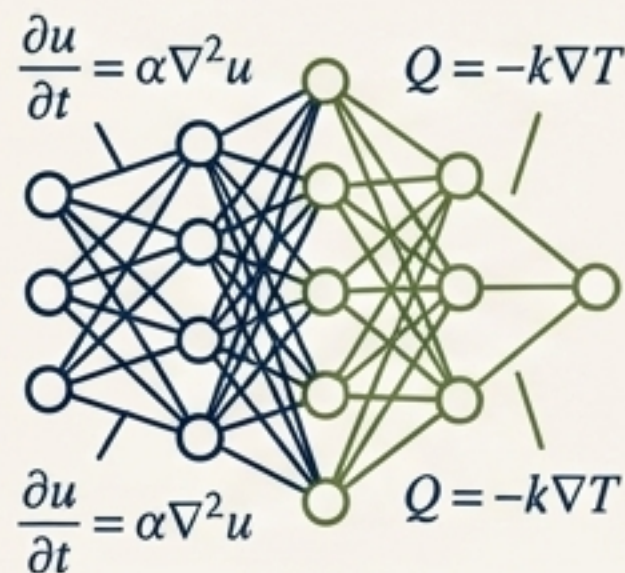


Konkretni primeri uporabe

Znanost

Formulacija problemov

Fizikalno podprte nevronske mreže (PINNs) vključujejo fizikalne zakone (npr. o prevajanju toplote) neposredno v model, kar omogoča natančnejše napovedi z manj podatki.



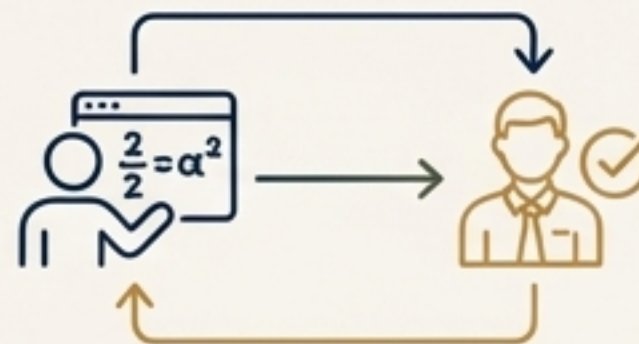
Inženirstvo

Reševanje problemov

Sistemi z več agenti, kjer en agent rešuje matematične probleme, drugi pa preverja rešitve, kar posnema realni proces in izboljšuje natančnost.

Razložljivost

MetaGPT integrira standardne operative postopke (SOP) v delovne tokove agentov, kar omogoča sledljivost in zmanjšuje napake pri razvoju programske opreme.



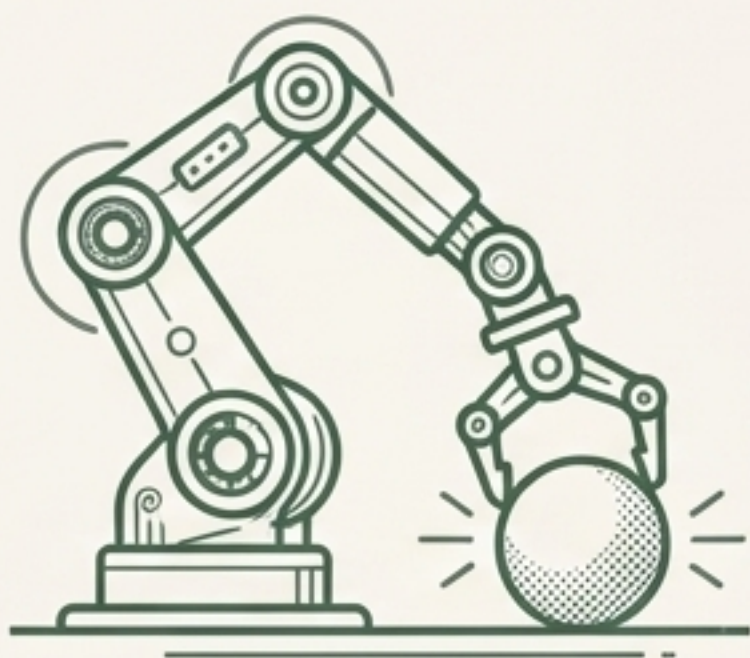
Družba

Razvoj hipotez

Simulacija virtualnih učilnic z agenti, ki predstavljajo učence in učitelje, za odkrivanje novih pedagoških načel.

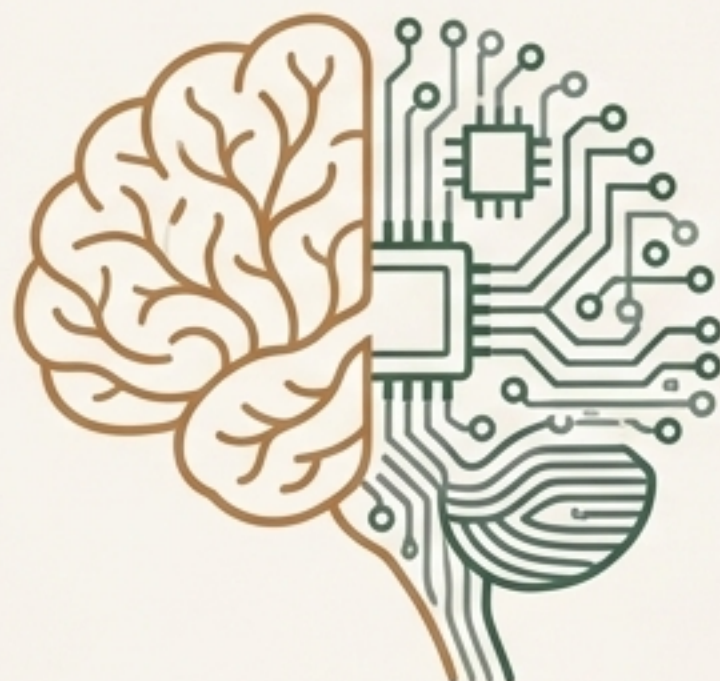


Pogled v prihodnost: Odprte raziskovalne smeri



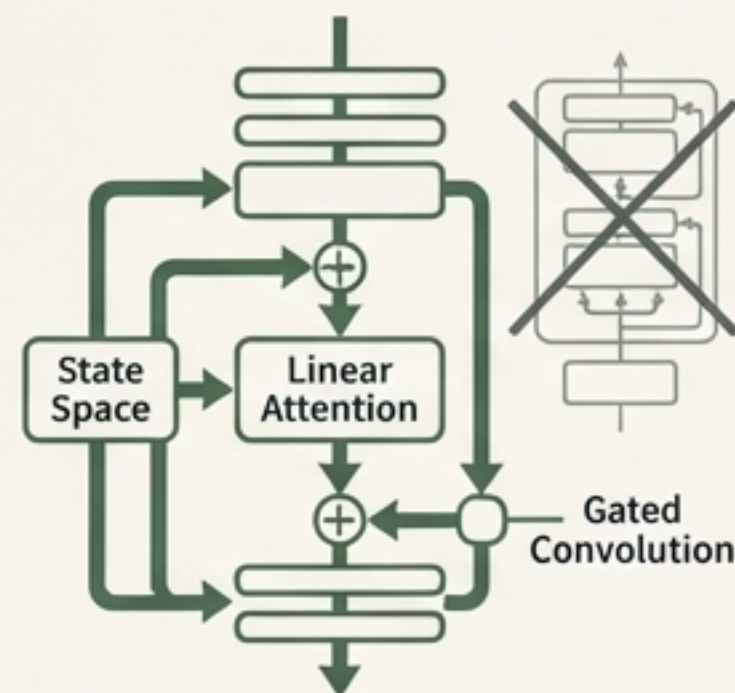
Utelešena UI (Embodied AI)

Razvoj modelov, ki se učijo in delujejo v fizičnem svetu. Izzivi vključujejo premoščanje vrzeli med simulacijo in resničnostjo ter učinkovito integracijo človeških demonstracij.



Možganom podobna UI (Brain-like AI)

Interdisciplinarno povezovanje z nevroznanostjo za razvoj modelov, ki posnemajo učinkovitost, prilagodljivost in vzporedno procesiranje človeških možganov.



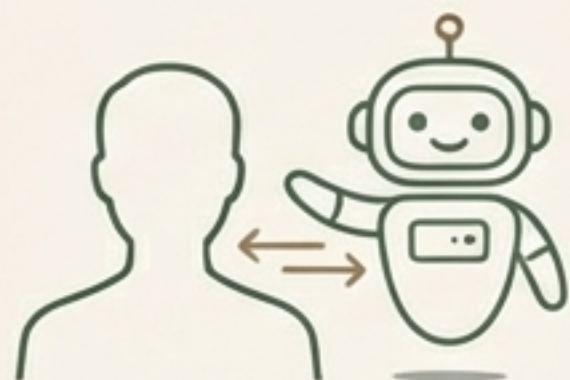
Ne-transformatorski temeljni modeli (Non-transformer Foundation Models)

Raziskovanje alternativnih arhitektur (npr. Hyena, Mamba, RetNet), ki ponujajo linearno skaliranje in večjo učinkovitost pri obdelavi dolgih zaporedij.

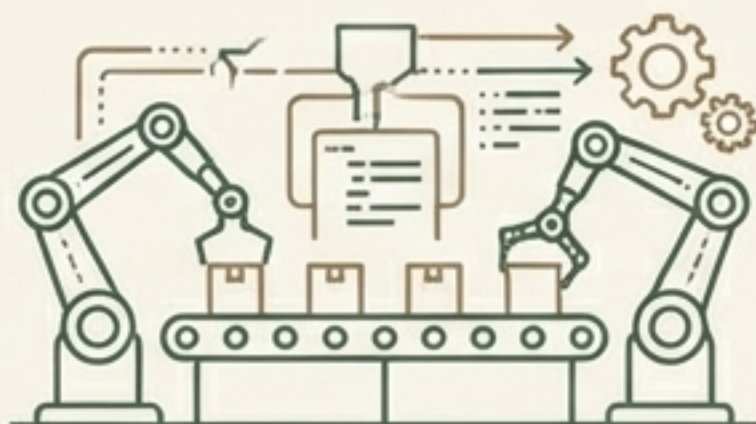
Pogled v prihodnost: Nastajajoče aplikacije



- **Omni-modalni sistemi:** Zlivanje podatkov iz različnih modalitet (besedilo, slike, zvok, senzorični signali) za celostno razumevanje fizičnega sveta, ključno za avtonomna vozila in digitalne dvojčke.



- **Personalizirani in prilagodljivi kolaborativni sistemi:** Sodelovanje velikih splošnih in malih specializiranih modelov za napredne inteligentne asistente, servisne robote in prilagojeno izobraževanje.



- **Napredna avtomatizacija:** Uporaba so-razvijajočih se modelov v avtonomni vožnji in industrijski proizvodnji za večjo robustnost in prilagodljivost v dinamičnih in nepredvidljivih okoljih.

Pot naprej: Ekosistem namesto monolita

Prihodnost UI presega monolitne modele.
Temelji na medsebojno povezanem ekosistemu,
ki je **obogaten z znanjem**, **sodelovalen**
in se **nenehno so-razvija**.

Ta načrt ponuja pot k izgradnji sistemov UI, ki niso le zmogljivejši,
temveč tudi bolj prilagodljivi, učinkoviti in usklajeni s kompleksnimi
potrebami znanosti, inženirstva in družbe.