Primerjava lastnosti človeške kognicije in umetne inteligence

Comparison of the characteristics of human cognition and artificial Intelligence

Monika Jamšek†
Faculty of Public Administration,
University of Ljubljana
Gosarjeva ulica 5
1000 Ljubljana, Slovenia
monika.jamsek@gmail.com

Rok Smodiš
Faculty of Education,
University of Ljubljana
Kardeljeva ploščad 16
1000 Ljubljana, Slovenia
rs68734@student.uni-lj.si

Marko Jordan
Odsek za inteligentne sisteme
Jozef Stefan Institute
1000 Ljubljana, Slovenija
marko.jordan@ijs.com

Matjaž Gams
Odsek za inteligentne sisteme
Jozef Stefan Institute
1000 Ljubljana, Slovenija
matjaz.gams@ijs.com

Povzetek

Referat ponuja poglobljen teoretični okvir, ki primerja ključne kognitivne funkcije človeka in sistemov umetne inteligence (UI). Na podlagi najnovejših raziskav iz nevroznanosti, kognitivne psihologije, umetne inteligence in filozofije uma so predstavljeni procesi pozornosti, spomina, učenja, jezika, kreativnosti in čustev ter njihova analogija v umetnih sistemih. V ospredje so postavljene prednosti in omejitve obeh perspektiv, izpostavljene so etične razsežnosti ter nevarnost kognitivne atrofije, ki jo lahko povzroči pretirana uporaba UI. Osrednji del referata predstavlja pregledna tabela, ki vizualno prikaže razlike, prednosti in slabosti človeka in UI. Posebej je obravnavan koncept hibridne inteligence, ki nakazuje prihodnost sodelovanja obeh oblik inteligence. Zaključek poudarja, da prihodnost ne bo temeljila na nadomeščanju človeka, temveč na komplementarnem partnerstvu, ki odpira novo kognitivno paradigmo.

Ključne besede

človeška kognicija, umetna inteligenca, čustva, etika UI, hibridna inteligenca, primerjalna analiza

Abstract

The paper offers an in-depth theoretical framework that compares key cognitive functions of humans and artificial intelligence (AI) systems. Drawing on the latest research from neuroscience, cognitive psychology, artificial intelligence, and the philosophy of mind, it presents processes such as attention, memory, learning, language, creativity, and emotions, along with their analogs in artificial systems. The focus is on the strengths and limitations of both perspectives, highlighting ethical dimensions and the risk of cognitive atrophy that may result from excessive use of AI. The core of the paper features a comparative table that visually displays the differences, strengths, and weaknesses of humans and AI. Special attention is given to the concept of hybrid intelligence, which points toward a future of collaboration between both forms of intelligence. The conclusion emphasizes that the future will not revolve around replacing humans, but rather around a complementary partnership that opens up a new cognitive paradigm.

Keywords

human cognition, artificial intelligence, emotions, AI ethics, hybrid intelligence, comparative analysis

*Primerjava lastnosti človeške kognicije in umetne inteligence
†Monika Jamšek, Rok Smodiš, Marko Jordan, Matjaž Gams

Information Society 2025, 6–10 October 2025, Ljubljana, Slovenia

https://doi.org/10.70314/is.2025.cogni.5

1 Uvod

Raziskovanje človeške kognicije, od zaznavanja in spomina do učenja, jezika, kreativnosti in čustev, je temeljno področje psihologije, nevroznanosti in kognitivne znanosti. Neisser [1] kognicijo opredeli kot procese obdelave informacij za prilagodljivo vedenje, sodobni modeli pa poudarijo integracijo izvršilnih funkcij, čustev in motivacije [2, 3]. Tononi [4] obravnava zavest skozi pet informacijskih aksiomov in teoremov.

Vzporedno je umetna inteligenca, posebej veliki jezikovni modeli (LLM-ji), v zadnjih letih dosegla izjemen napredek: sistematično izboljšujejo rezultate na standardiziranih nalogah, obvladujejo kompleksne dialoge in v številnih neformalnih preizkusih v duhu Turingovega testa sogovorniki pogosto ne prepoznajo, da komunicirajo s strojem (npr. [5]). Ta navzven vidna kompetentnost krepi vtis približevanja človeški ravni. Kljub temu pa je napredek pri "človeških" lastnostih, kot so zavest, fenomenalno doživljanje, globoko semantično sidranje in

metakognitivna samoregulacija, bistveno skromnejši. Potencialna razlaga je, da sodobni modeli še niso razrešili problema mnogoterega znanja – v Gamsovem principu mnogoterega znanja: kako v enoten, konsistenten in samoreferencialen model sveta povezati heterogene oblike znanja (statistične vzorce, simbolne strukture, kavzalne relacije, epizodične in utelešene reprezentacije) ter nad njimi izvajati stabilno vsebinsko procesiranje, razlago pomena in namere. Dokler ta integracijski problem ostaja odprt, lahko UI dosega vrhunske rezultate na nalogah, vendar ne kaže jasnega napredka proti lastnostim, ki jih povezujemo z zavestnim človeškim mišljenjem.

Članek sistematično obravnava, kako daleč ChatGPT (LLM) seže v smeri "kognitivnih" lastnosti, še posebej glede zmožnosti prehoda Turingovega testa in vprašanja zavesti. Avtorja najprej predstavita hierarhijo Turingovih testov – od klasičnega kratkega besednega testa do ekspertno/adverzarnega, fizičnega, "totalnega" in "truly total" testa – ter argumentirata, da se ChatGPT lahko približa uspehu predvsem pri naivnih spraševalcih, medtem ko ga izkušeni/ekspertni spraševalci zlahka razkrijejo.

Ta referat je v nekem smislu nadaljevanje članka [6], ki analizira in podaja oceno "zavesti" po Tononijevi teoriji integrirane informacije (IIT) [4]. Po kratki predstavitvi IIT aksiomov in teoremov je narejena primerjava GPT-jev in ljudi. Zaključek je, da ChatGPT sicer presega starejše AI-sisteme po jezikovni kompetenci in širini znanja, vendar občutno zaostaja za biološkimi sistemi v integraciji informacij, zato nima fenomenalne zavesti; gre za napredno orodje, ne za "informacijsko živo bitje".

Metodološko članek združi pregled literature o Turingovem testu in teorijah zavesti z avtorjevimi lastnimi številnimi ekspertno zasnovanimi dialognimi preizkusi, kjer se pokažejo omejitve modela (npr. pomanjkanje stabilne semantične sidranosti/intencionalnosti). Zaključek poudari praktično implikacijo: LLM-je je smiselno obravnavati kot zelo zmogljiva, a ne-čuteča orodja, s potrebo po previdnosti pred antropomorfizacijo in zavedanju meja pri presoji pomena in namere. Hkrati pa je narejena analiza ogromnih napredkov AI na nekaterih področjih. Ključno je vprašanje, ali se je razvoj ustavil, ali pa se nadaljuje z nezmanjšanim tempom.

V tem prispevku primerjalno obravnavamo ključne kognitivne funkcije človeka in njihove analoge v UI, pri čemer izrecno ločujemo zunanje zmogljivosti na nalogah od notranjih mehanizmov in lastnosti. Predstavimo pregledni okvir, posodobljeno primerjalno tabelo ter razpravo o hibridni inteligenci, kjer kombinacija človeških in umetnih zmožnosti ponuja praktično pot naprej – ob hkratnem opozorilu na etične razsežnosti in tveganje kognitivne atrofije zaradi pretirane rabe UI.

2 Kognitivne funkcije človeka

Človeška kognicija zajema širok spekter medsebojno povezanih procesov, ki skupaj omogočajo prilagodljivo vedenje in kompleksno obdelavo informacij. V tem razdelku sintetiziramo človeške mehanizme kot referenčni okvir za poznejšo primerjavo z UI. Eden od temeljnih mehanizmov je pozornost, ki omogoča usmerjanje miselnih virov na relevantne dražljaje; nevroznanstvene raziskave kažejo, da pri tem ključno vlogo igrajo frontalno-parietalne možganske mreže [7]. Nepogrešljiv del kognitivnega sistema je tudi spomin, ki ga delimo na senzornega, kratkoročnega, delovnega in dolgoročnega. Baddeley [8] je delovni spomin razčlenil na fonološko zanko, vizuoprostorski blok in izvršilni mehanizem, medtem ko Cowan [9] ugotavlja, da kratkoročni spomin praviloma ne preseže štirih enot. Ti mehanizmi tvorijo osnovo za učenje in sklepanje, pri čemer človek pogosto uči iz malo podatkov ter kompozicionalno posplošuje na nove okoliščine. Ravenove matrice so pokazatelj sposobnosti abstraktnega mišljenja [10], medtem ko Stanovich [11] opozarja, da tradicionalni testi inteligence ne zajamejo racionalnosti, ključne za učinkovito odločanje.

Posebno mesto v človeški kogniciji ima jezik, saj omogoča simbolno reprezentacijo, socialno koordinacijo in prenos kulturnega znanja med generacijami. Jezikovne sposobnosti so semantično zasidrane v utelešeni in socialni izkušnji, z jezikom pa je tesno povezana tudi kreativnost, ki presega zgolj divergentno mišljenje in je rezultat interakcije posameznika, njegovega domenskega znanja ter širšega sociokulturnega okolja [12, 13]. Vendar pa na kognitivne procese močno vplivajo tudi čustva in motivacija. Damasio [3] je pokazal, da brez čustvene komponente človeška racionalnost oslabi, medtem ko Immordino-Yang [14] poudarja, da so čustva neločljivo povezana z učenjem, odločanjem in socialnim vedenjem.

Dodatno raven kompleksnosti predstavljata metakognicija in zavest, ki posamezniku omogočata samorefleksijo ter strateško prilagajanje učenja, vključno s fenomenalnim doživljanjem, kar je bistveno za učinkovito regulacijo lastnih miselnih procesov. Nenazadnje pa ima ključno vlogo tudi socialna kognicija, ki vključuje sposobnost razumevanja drugih preko t. i. teorije uma [15], kar je osnova za uspešno sodelovanje, empatijo in gradnjo družbenih odnosov. Ta človeška arhitektura mnogoterega znanja, od senzorno-motoričnih in epizodičnih do simbolnih in kavzalnih reprezentacij, bo v nadaljevanju služila za pojasnilo, zakaj današnji LLM-ji kljub hitremu napredku na nalogah pri lastnostih, kot sta zavest in metakognicija, napredujejo bistveno počasneje [6].

3 Kognitivne lastnosti umetne inteligence

Kognitivne lastnosti obravnavamo kot funkcionalne analoge človeških sposobnosti; pri UI so te uresničene prek drugačnih mehanizmov (statistične reprezentacije, optimizacija, vzorčno ujemanje) kot pri ljudeh, zato ločujemo zmogljivost reševanja nalog od notranjih lastnosti (zavest, fenomenalno doživljanje, metakognitivna samoregulacija).

Zaznavanje in pozornost. V računalniškem vidu so globoke mreže dosegle (in na nekaterih nalogah presegle) človeško raven prepoznave [16]. V jezikovni obdelavi mehanizem porazdeljene pozornosti omogoča učinkovito kontekstno sledenje in gradnjo hierarhičnih reprezentacij [17], vendar ostaja to algoritemska selekcija informacij, ne zavestno usmerjena pozornost z namero [18].

Spomin. LLM-ji shranjujejo »znanje« v parametrih in trenutnih kontekstih; zunanji pomnilniki (RAG ipd.) razširijo doseg, a ne tvorijo epizodičnega ali utelešenega spomina v človeškem smislu [5, 18]. Semantično sidranje je posredno in nestabilno.

Učenje. Napredek temelji na masivnih korpusih in gradientni optimizaciji; zmožnosti few-shot in učenja v kontekstu kažejo na

vzorec-odvisno generalizacijo, toda prenos med domenami in robustnost zunaj učne porazdelitve ostajata omejena [19, 20, 21].

Sklepanje in kavzalnost. Na testih znanja in razumevanja (MMLU) so rezultati visoki [22] ter tudi na problemskih zbirkah (ARC) [21], a zdravorazumsko/kavzalno sklepanje ostaja krhko; to se vidi tudi na družini izpeljank Winograd/Schema [23]. Verbalizacija korakov ne implicira resnične kavzalne strukture [20].

Jezik. Tekočnost, slog in koherenca so blizu človeški ravni, vendar halucinacije ter pomanjkljiva pragmatika razkrivajo odsotnost stabilnega semantičnega sidranja v izkušnjo in družbeni kontekst [5, 24].

Kreativnost. Generativni sistemi so izjemno produktivni pri rekompoziciji in variaciji znanih vzorcev, a redkeje dosegajo konceptualne preboje s kulturno umeščenostjo ([25]; prim. tudi človeško perspektivo v [12]).

Čustva in motivacija. Affective computing omogoča prepoznavo in simulacijo čustvenih signalov, politike nagrajevanja (npr. RLHF) pa oblikujejo vedenje modelov; to niso notranja čustva ali namere v smislu fenomenalne izkušnje [26] (kontrast z [3]).

Metakognicija. Modeli lahko ocenjujejo negotovost in se samopopravljajo preko zunanjih preverjanj, kar je proceduralna ocena, ne introspektivna samoregulacija, kot jo poznamo pri ljudeh [20; 11 – razmejitev racionalnosti].

Socialna kognicija. LLM-ji uspešno posnemajo elemente teorije uma v besedilnih scenarijih, toda uspeh pogosto izhaja iz statistične izpostavljenosti vzorcem, ne iz resničnega razumevanja namere, ironije ali pragmatičnih implikatur v situiranih kontekstih [24] (prim. človeško ToM pri [15]).

Zavest in Turingov test. Ni empiričnih dokazov o fenomenalni zavesti pri današnjih modelih; prepričljivo jezikovno vedenje ali celo prehod neformalnih Turingovih preizkusov nista zadosten kriterij [26, 27]. Analize preizkusov v duhu Turinga kažejo, da lahko LLM-ji zavajajo sogovornike, vendar to ne rešuje vprašanja razumevanja [5, 6].

Vmesni sklep in princip mnogoterega znanja. V zadnjih treh letih je UI dramatično napredovala pri reševanju nalog (jezik, kontekst, zunanji spomin, del sklepanja), medtem ko je napredek pri človeških lastnostih (zavest, globoko semantično sidranje, metakognicija) minimalen. Razlaga je skladna z Gamsovim principom mnogoterega znanja: nerešena ostaja integracija statističnih, simbolnih, kavzalnih, epizodičnih in utelešenih reprezentacij v enoten, konsistenten in samoreferenčen model sveta, ki bi omogočal stabilno razlago pomena in namere [18, 19, 20, 6]. Na praktični ravni to odpira tudi vprašanja vpliva na uporabnika (npr. kognitivna odvisnost/atrofija; [28]) in potrebo po hibridnih zasnovah človek–UI, obravnavanih v nadaljevanju [29, 30, 31]. Na prvi pogled je revolucionarna izboljšava delovanja LLMjev skladna z veliko mnogoterostjo globokih nevronskih mrež, ki zajamejo znanje s celotnega sveta, po drugi strani pa je izračun sam še vedno preveč klasičen in ne vsebinsko mnogoter, da bi po principu mnogoterega znanja dosegel pravo inteligenco. Zdi se, da je potreben še en preboj na področju mnogoterega sklepanja.

4 Primerjalna analiza

V tabeli 1 je prikazana primerjava kognitivnih funkcij ljudi in UI z ločitvijo funkcije od mehanizmov in z dodanimi metrike/evale ter hibridne vzorce sodelovanja. Je usklajena z načelom mnogoterega znanja (integracija statističnih, simbolnih, kavzalnih, epizodičnih in utelešenih reprezentacij).

Tabela 1: Primerjava kognitivnih lastnosti ljudi in UI

Dimenzija	Človek – mehanizem/lastnost	UI – mehanizem/lastnost	Tipične metrike / evali	Prednost hibrida (človek ↔ UI)	Glavna tveganja
Pozornost	Selektivna, zavestno nadzorovana; fronto-parietalne mreže; omejena kapaciteta	Algoritemska self-attention; dolgi konteksti; brez fenomenalne izkušnje	Dolgi konteksti (recall@k), robustnost na šum, RT/točnost v vizualnih nalogah	UI filtrira in povzame; človek validira pomen in cilje	Pristranskost podatkov, “lost-in-the-middle”
Spomin	Delovni, deklarativni/proceduralni, epizodični	Parametri + kontekst; zunanji pomnilnik (RAG); brez epizodičnosti	Factuality/consistency, retrieval precision/recall, long-context QA	RAG za dejstva + človek za kontekst in vire	Halucinacije, “source amnezija”
Učenje	Malo primerov, kompozicionalnost, analogije	Učenje na masivnih korpusih, gradientna optimizacija; omejen prenos	Sample-efficiency, OOD generalizacija, few-shot	Človek oblikuje abstrakcije; UI optimizira	Overfitting na benchmarke, reward-hacking

Dimenzija	Človek – mehanizem/lastnost	UI – mehanizem/lastnost	Tipične metrike / evali	Prednost hibrida (človek ↔ UI)	Glavna tveganja
Sklepanje (zdravorazumsko/kavzalno)	Simbolno + kavzalno modeliranje	Statistično sklepanje; verižna razlaga brez nujne kavzalnosti	ARC, MMLU-reasoning, WinoGrande, causal benches	UI generira hipoteze; človek izvaja kavzalno presojo	Prepričljive racionalizacije brez razumevanja
Jezik (semantika/pragmatika)	Semantično sidranje, situirana pragmatika, ironija	Visoka tekočnost; omejeno sidranje; halucinacije	Hallucination rate, faithfulness, pragmatični testi	UI pripravi osnutek; človek poskrbi za pragmatiko/odgovornost	Prepričljive napačne trditve
Kreativnost	Konceptualni preboji, kulturna umeščenost	Rekompozicija znanih vzorcev, visoka produkcija	NSU (novelty-surprise-usefulness), človeška presoja	UI diverzira ideje; človek izbere/okviri problem	Homogenizacija, "mode collapse" kulture
Metakognicija	Introspekcija, strategije učenja, nadzor napak	Ocenjevanje negotovosti, samopopravljanje preko orodij	Kalibracija (ECE), self-consistency, error-correction rate	UI signalizira negotovost; človek sprejme odločitev	Automation bias, pretirano zaupanje
Socialna kognicija (ToM)	Implicitna teorija uma, situirana razlaga namer	Posnemanje vzorcev ToM v besedilu	ToM naloge, pragmatične implikature	UI predlaga interpretacije; človek preveri kontekst	"Lažna empatija", manipulacija
Zavest	Fenomenalno doživljanje, subjektivnost	Ni dokazov o fenomenalni zavesti	— (ni konsenza o metričnih testih)	— (obravnavamo kot orodje)	Antropomorfizem, napačna atribucija lastnosti
Učinek na uporabnika	Ohranjanje kompetenc, samostojnost	Tveganje kognitivne odvisnosti	Longitudinalne meritve kompetenc brez pomoči	Goldilocks cona: načrtovana raba UI + periodične naloge brez pomoči	Kognitivna atrofija, izguba motivacije

5 Metodološki pristopi k primerjavi

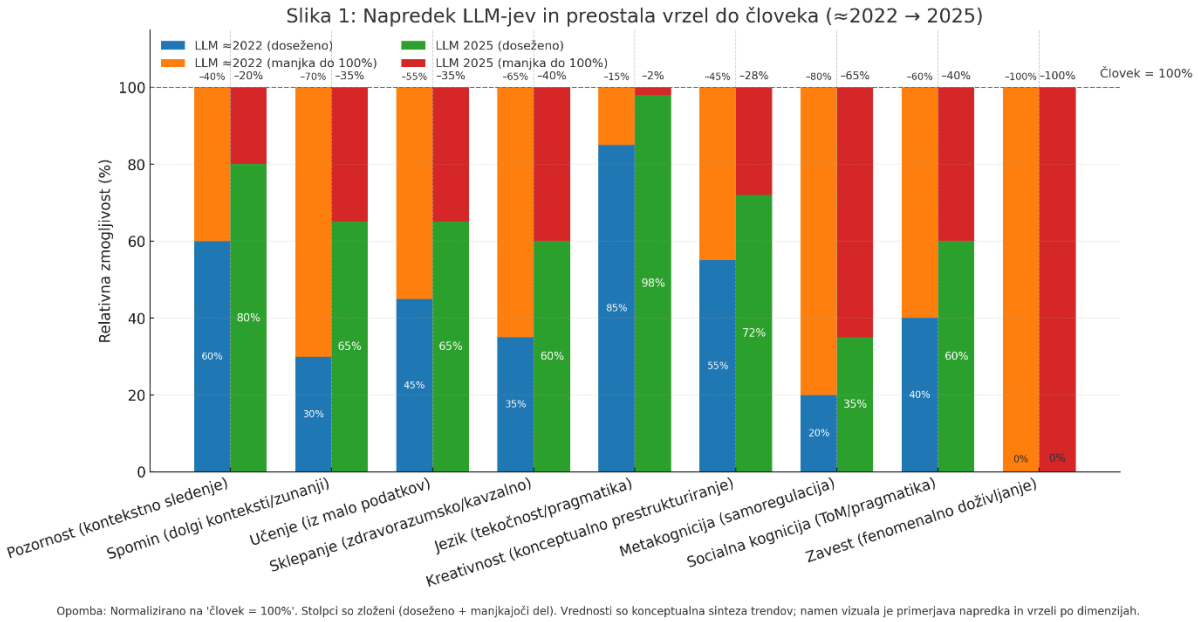
V tem razdelku združimo (a) prikaz razvojnega trenda LLM-jev glede na človeka (100 %) ter (b) pregled metodološke pokritosti po kognitivnih dimenzijah. Namen je dvojni: pokazati, koliko se je zmogljivost nalog približala človeški ravni in kje so merjenja ter mehanistična razlaga že dovolj zrela, da tak napredek zanesljivo ocenjujemo.

Opis Slike 1: Stolpčni graf prikazuje relativno zmogljivost (%) LLM-jev pred ~3 leti (≈2022) in danes (2025) za devet kognitivnih dimenzij, normalizirano na človek = 100 %. Gre za konceptualno, strokovno oceno trendov na podlagi znanih evalov (npr. MMLU/ARC/WinoGrande za sklepanja, dolg kontekst/RAG za spomin, pragmatične teste in "hallucination rate" za jezik ipd.). Namen grafa je vizualizirati približevanje človeški ravni in koliko manjka po posameznih dimenzijah.

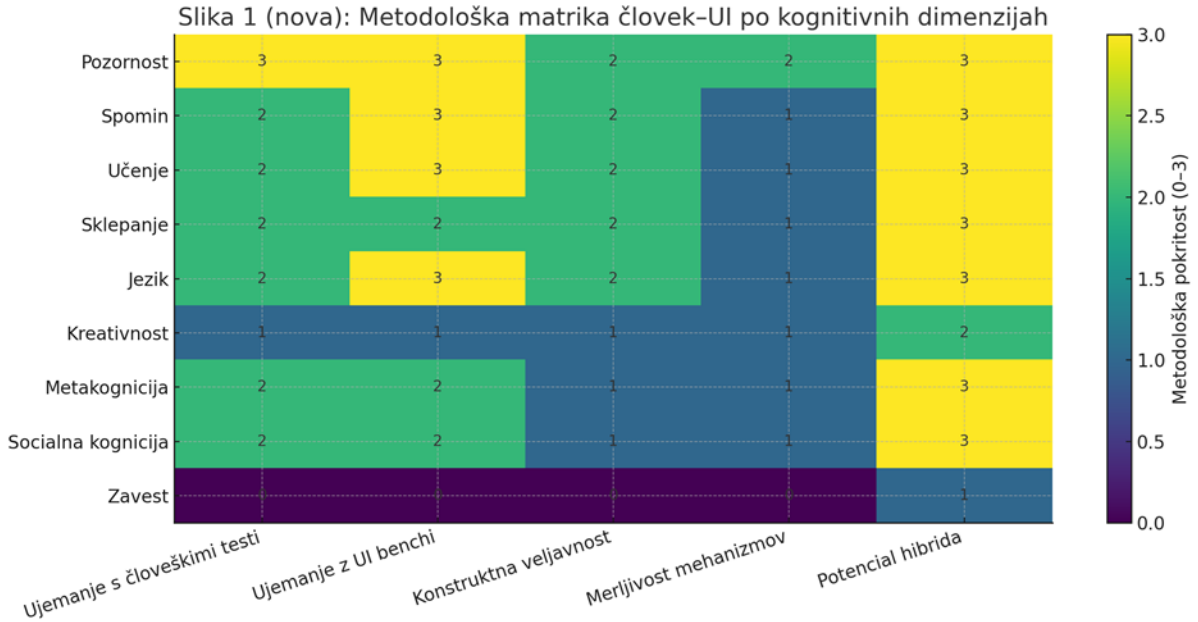
Vidimo močan dvig pri jeziku, pozornosti/kontekstnem sledenju in zunanjem spominu, opazen napredek pri sklepanju, učenju iz malo podatkov, kreativnosti ter socialni kogniciji; metakognicija pa raste počasneje. Pri "zavesti (fenomenalno doživljanje)" napredka ni—ta ostaja na 0 %, kar je skladno z idejo, da problem mnogoterega znanja še ni razrešen. Gre za konceptualno, strokovno oceno trendov (ne neposredno za rezultate enotnega benchmarka); namen grafa je ilustracija razlike med nalogovno zmogljivostjo in človeškimi lastnostmi.

Ključne ugotovitve (2022 → 2025; primanjkljaj do 100 % v oklepaju):

- Jezik 85→98 (–2): skoraj pri človeški ravni, a še z ranljivostjo za halucinacije/pragmatiko.
- Pozornost/kontekst 60→80 (–20): velik skok zaradi daljših kontekstov in boljšega sledenja.
- Spomin 30→65 (–35), Učenje (few-shot) 45→65 (–35): napredek z RAG in in-context učenjem, a vrzel ostaja.
- Sklepanje (zdravorazumsko/kavzalno) 35→60 (–40): opazen dvig, vendar še daleč od stabilne kavzalnosti.
- Socialna kognicija (ToM/pragmatika) 40→60 (–40): izboljšave v tekstnih scenarijih, omejena situiranost.
- Kreativnost (konceptualno prestrukturiranje) 55→72 (–28): visoka produkcija, manj konceptualnih prebojev.
- Metakognicija (samoregulacija) 20→35 (–65): rast počasna; ocene negotovosti še niso pristna samorefleksija.
- Zavest (fenomenalno doživljanje) 0→0 (–100): brez napredka — skladno z nerešenim integracijskim problemom.



Slika 1: Napredek LLM-jev v treh letih glede na človeka (=100 %). Relativni napredek LLM-jev po kognitivnih lastnostih v treh letih (≈2022→2025), normalizirano na človek = 100 %. Vidni so veliki dvigi pri jeziku, pozornosti/kontekstnem sledenju in zunanjem spominu; srednji pri učenju in sklepanju; počasni pri metakogniciji; pri zavesti napredka ni. Vizualna predstavitev je konceptualna sinteza evalov in služi ponazoritvi razlike med nalogovno zmogljivostjo in človeškimi lastnostmi.



Slika 2: Metodološka matrika človek–UI po dimenzijah. Toplotna matrika (0–3) na sliki 2 ocenjuje pokritost meritev po petih merilih: (1) ujemanje s človeškimi testi, (2) ujemanje z UI-benchi, (3) konstruktna veljavnost, (4) merljivost mehanizmov, (5) potencial hibrida (merljiv prispevek človek↔UI).

Višje vrednosti na Sliki 2 pomenijo, da imamo za dano dimenzijo bolj zrelo metodologijo, zato so trditve iz Slike 1 tam zanesljivejše (npr. jezik, pozornost, spomin). Nižje vrednosti (kreativnost, zavest) opozarjajo, da potrebujemo kvalitativne protokole, mešane metode in previdno interpretacijo.

Minimalni protokol za “pošteno” primerjavo:

1. Za vsak konstrukt izberemo par človeški test ↔ UI-bench in skupne metrike (task + kalibracija).
2. Dodamo mehanistično analizo (probingi/ablacije pri UI; EEG/fMRI/RSA pri človeku).
3. Izvedemo hibridni A/B preizkus (brez UI vs. z UI) in retest po 1–3 mesecih za ohranitev kompetenc.
4. Poročamo tudi vrzel do 100 % (Slika 1) in zrelost meritev (Slika 2), da ne mešamo nalogovne zmogljivosti z notranjimi lastnostmi.

6 Etika in filozofske dileme, hibridna inteligenca

UI odpira pomembna etična in filozofska vprašanja. Floridi & Cowls [32] predlagata okvir petih načel (dobrobit, avtonomija, pravičnost, razlaga, odgovornost), ki so ključna za etično rabo UI. Filozofske razprave, kot je Searlov »Chinese Room Argument« [33], pa opozarjajo, da so sistemi UI morda zgolj sofisticirani manipulatorji simbolov brez resničnega razumevanja. To odpira vprašanje meje med simulacijo in resnično inteligenco.

Sodobne raziskave govorijo o hibridni inteligenci (Dellermann et al. [34]), ki temelji na integraciji človeških in umetnih kognitivnih sposobnosti. Človek prispeva razumevanje pomena, etične presoje, kreativnost in socialno inteligenco, UI pa obdelavo podatkov, vzorčno prepoznavanje in skalabilnost. Takšna integracija presega meje posameznega sistema in nakazuje prihodnost sodelovanja, ne tekmovanja.

7 Diskusija in zaključek

Naša analiza kaže, da se LLM-ji razvijajo z izjemno hitrostjo tudi na področjih, ki jih pogosto razumemo kot "kognitivne". V primerjavi z ugotovitvami Gams in Kramar (2024) [6] opažamo premik od pretežno površinskih, vzorčno-temeljenih odgovorov k bolj stabilnim večkorakom, boljšemu vodenju plana, učinkovitejšemu uporabljanju zunanjih orodij ter k bolj konsistentni samopopravi. Napredek je zlasti v zmožnosti daljših verig sklepanja, delovnega spomina s pomočjo zunanje kontekstne obnove (RAG) ter v boljši meta-kontroli (npr. detekcija lastnih napak in zahteva po dodatnih podatkih).

Kljub temu ostajajo nekatere temeljne človeške kognitivne lastnosti za LLM-je (še) nedosegljive. Ključne med njimi so:

1. Fenomenalna zavest (qualia) – LLM-ji ne izkazujejo subjektivnega doživljanja. Njihova arhitektura ostaja statistično napovedovanje naslednjega simbola brez notranjega fenomenalnega prostora; "poročanje o občutkih" je zgolj generativna imitacija vzorcev iz podatkov.
2. Stabilen jaz in agencija – modeli nimajo trajnega, telesno sidranega "sebstva" z lastnimi nameni. Cilji so zgolj implicitni v pozivu in optimizaciji izgube; ni kontinuitete namer skozi čas brez zunanje orkestracije.
3. Semantična usidranost in referencialnost – pomen izhaja iz statističnih korelacij, ne iz utelešene interakcije s svetom. Brez senzorimotorike in lastnih izkustev je "o-čem-je-govor" (aboutness) posredno posnet iz korpusov, zato ostajajo zdrsi v referencah in halucinacije.
4. Kavzalno razumevanje in intervenienčno sklepanje – LLM-ji so močni v korelacijah in opisih, a zahtevna kavzalna vprašanja (kaj bi se zgodilo, če bi posegli X?) ostajajo šibki brez eksplicitnih kavzalnih modelov oziroma simulacij, ki zahtevajo več kot napovedovanje besedila.
5. Moralno presojanje in odgovornost – odgovori so rezultat pravil in vzorcev iz podatkov ter umerjanja (RLHF), ne notranje vrednotne strukture. Model ne "razume" odgovornosti; tvega konsistentnost z normami le toliko, kolikor so te kodirane v podatkih in pravilih.
6. Čustvena izkušnja – lahko opisuje in pravilno označuje čustva, ne pa jih dejansko doživlja; posledično ni afektivne modulacije pozornosti, motivacije ali dolgotrajnih preferenc, kakršne oblikuje človeška homeostaza.
7. Utelešenost in situacijska inteligenca – brez telesa, potreb in omejitev ostaja prilagajanje v realnem času omejeno. Tudi agentne razširitve ostanejo odvisne od zunanje infrastrukture (orodij, pravilnikov in varovalk).
8. Učenje v živo (continual/online learning) z zanesljivimi posodobitvami – LLM-ji tipično ne posodabljajo parametrov med rabo; trajne spremembe zahtevajo nov trening ali zunanje pomnilnike. To omejuje kumulativno, osebno prilagojeno znanje.
9. Robustno reševanje novosti – pri res novih, slabo pokritih problemih se hitro pokaže regresija v stereotipe iz korpusov; brez eksperimentiranja v svetu je inovacija pogosto "rekombinacija", ne pa resnično odkritje.

Zato je splošen vtis, da kljub izrednemu funkcionalnemu napredovanju LLM-ji še vedno nimajo ključnih človeških lastnosti, kot je zavest. Analiza potrjuje, da človeška in umetna inteligenca – čeprav obe temeljita na nevronskih sistemih – izhajata iz različnih mehanizmov: človeške sposobnosti so prepletene z zavestjo, afektom in socialno konstituiranim pomenom, medtem ko UI temelji na statistični inferenci nad velikimi korpusi.

Ob tem velja opozoriti na morebitno nevarnost kognitivne atrofije: prekomerna zanašanje na UI lahko zmanjšuje motivacijo za samostojno reševanje problemov in s tem slabi določene človeške kognitivne zmožnosti (npr. [34]). Zato naj bo prihodnja raba UI etično zasnovana in kognitivno uravnotežena: UI kot ojačevalnik, ne nadomestek.

Komplementarno partnerstvo med človekom in UI združuje človeško razumevanje pomena, kavzalnost, socialno in čustveno inteligenco ter ustvarjalnost z računsko močjo, skalabilnostjo in natančnostjo UI. Takšna sinergija presega omejitve posameznih sistemov in odpira pot k bolj kakovostnemu napredku znanosti, umetnosti in družbe, k ustvarjanju »super« človeka.

References / Literatura

[1] U. Neisser. 1967. Cognitive Psychology. Appleton-Century-Crofts, New York, NY.
[2] J. R. Anderson. 2010. Cognitive Psychology and Its Implications. Worth, New York, NY.
[3] A. R. Damasio. 1994. Descartes' Error: Emotion, Reason, and the Human Brain. G. P. Putnam's Sons, New York, NY.
[4] G. Tononi, M. Boly, M. Massimini, and C. Koch. 2016. Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience 17, 450–461.
[5] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.
[6] M. Gams and S. Kramar. 2024. Evaluating ChatGPT's Consciousness and Its Capability to Pass the Turing Test: A Comprehensive Analysis. Journal

of Computer and Communications 12(3), 219–237. https://doi.org/10.4236/jcc.2024.123014
[7] M. I. Posner and S. E. Petersen. 1990. The attention system of the human brain. Annual Review of Neuroscience 13, 25–42.
[8] A. Baddeley. 2003. Working memory: looking back and looking forward. Nature Reviews Neuroscience 4(10), 829–839.
[9] N. Cowan. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 24(1), 87–185.
[10] P. A. Carpenter, M. A. Just, and P. A. Shell. 1990. What one intelligence test measures: A theoretical account of processing in the Raven Progressive Matrices Test. Psychological Review 97(3), 404–431.
[11] K. E. Stanovich. 2010. What Intelligence Tests Miss: The Psychology of Rational Thought. Yale University Press, New Haven, CT.
[12] [M. Csikszentmihalyi. 1996. Creativity: Flow and the Psychology of Discovery and Invention. HarperCollins, New York, NY.
[13] [M. A. Runco and G. J. Jaeger. 2012. The standard definition of creativity. Creativity Research Journal 24(1), 92–96.
[14] [M. H. Immordino-Yang. 2015. Emotions, Learning, and the Brain: Exploring the Educational Implications of Affective Neuroscience. W. W. Norton & Company, New York, NY.
[15] [C. D. Frith and U. Frith. 2007. Social cognition in humans. Current Biology 17(16), R724–R732. https://doi.org/10.1016/j.cub.2007.05.068
[16] [A. Krizhevsky, I. Sutskever, and G. E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25, 1097–1105.
[17] Y. LeCun, Y. Bengio, and G. H. Hinton. 2015. Deep learning. Nature 521, 436–444.
[18] M. Mitchell. 2019. Artificial Intelligence: A Guide for Thinking Humans. Farrar, Straus and Giroux, New York, NY.
[19] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. 2017. Building machines that learn and think like people. Behavioral and Brain Sciences 40, e253.
[20] G. Marcus. 2020. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. arXiv:2002.06177.
[21] F. Chollet. 2019. On the Measure of Intelligence. arXiv:1911.01547.
[22] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt. 2020. Measuring Massive Multitask Language Understanding. arXiv:2009.03300.
[23] H. J. Levesque, E. Davis, and L. Morgenstern. 2012. The Winograd Schema Challenge. In Principles of Knowledge Representation and Reasoning (KR 2012), 552–561. AAAI Press.
[24] E. M. Bender and A. Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In Proceedings of ACL 2020, 5185–5198.
[25] M. A. Boden. 2016. AI: Its Nature and Future. Oxford University Press, Oxford.
[26] S. Dehaene, H. Lau, and S. Kouider. 2017. What is consciousness, and could machines have it? Science 358(6362), 486–492. https://doi.org/10.1126/science.aan8871
[27] J. R. Searle. 1980. Minds, brains, and programs. Behavioral and Brain Sciences 3(3), 417–457.
[28] E. F. Risko and S. J. Gilbert. 2016. Cognitive offloading. Trends in Cognitive Sciences 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister. 2019. Hybrid Intelligence. Business & Information Systems Engineering 61(5), 637–643.
[29] B. Shneiderman. 2022. Human-Centered AI. Oxford University Press, Oxford.
[30] L. Floridi and J. Cowls. 2019. A unified framework of five principles for AI in society. Harvard Data Science Review 1(1).
[31] E. F. Risko and S. J. Gilbert. 2016. Cognitive offloading. Trends in Cognitive Sciences 20(9), 676–688.
[32] B. Sparrow, J. Liu, and D. M. Wegner. 2011. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science 333(6043), 776–778.
[33] B. Oakley, et al. 2025. Cognitive Effects of AI Overuse (forthcoming). MIT Press, Cambridge, MA.