Razvoj in delovanje umetne inteligence

Marko Radovan[1], *Anton Gradišek*[2]

Povzetek

V tem besedilu se bomo na seznanili z razvojem in značilnostmi umetne inteligence (UI), ki je postala del našega vsakdanjika. Osredotočili se bomo na generativno UI, kot so ChatGPT in DALL-E, ki spontano prodirajo v izobraževanje in prinašajo priložnosti za individualizacijo učenja ter različne izzive. V prvem delu spoznamo osnove UI – sisteme, ki se učijo iz podatkov. Razlikujemo med ozko UI (današnji specializirani sistemi) in splošno UI (še ne obstaja). Sledil bo zgodovinski pregled od Turingovega testa (1950) do sodobnih transformatorjev. Nadaljevali bomo z generativno UI, ki ustvarja nove vsebine in deluje večmodalno. Razložili bomo, kako veliki jezikovni modeli (LLM) statistično napovedujejo besede, ne razumejo pa sveta vzročno, kar povzroča halucinacije. Opozorili bomo, da UI ni nepristranska in odraža pristranskosti učnih podatkov. Zaključili bomo s ključno nevarnostjo – kognitivnim dolgom, kjer prekomerna odvisnost od UI oslabi kritično mišljenje. Poudarili bomo, da mora UI ostati orodje za razširitev razmišljanja, učitelj pa nepogrešljiv vodja učnega procesa.

Ključne besede: generativna umetna inteligenca, kognitivni dolg, pedagoška nadgradnja, algoritemska pristranskost

The Development and Operation of Artificial Intelligence

Abstract

In this text, we will become acquainted with the development and characteristics of artificial intelligence (AI), which has become part of our everyday life. We will focus on generative AI, such as ChatGPT and DALL-E, which are spontaneously penetrating education and bringing opportunities for individualisation of learning as well as various challenges. In the first part, we will explore the basics of AI – systems that learn from data. We distinguish between narrow AI (today's specialised systems) and general AI (does not yet exist). This will be followed by a historical overview from the Turing test (1950) to contemporary transformers. We will continue with generative AI, which creates new content and operates multimodally. We will explain how large language models (LLMs)

1 Filozofska fakulteta, Univerza v Ljubljani, marko.radovan@ff.uni-lj.si

2 Inštitut Jožef Stefan, anton.gradisek@ijs.si

statistically predict words but do not understand the world causally, which causes hallucinations. We will point out that AI is not unbiased and reflects the biases of training data. We will conclude with a key danger – cognitive debt, where excessive reliance on AI weakens critical thinking. We will emphasise that AI must remain a tool for expanding thinking, whilst the teacher remains an indispensable leader of the learning process.

Keywords: generative artificial intelligence; cognitive debt; pedagogical enhancement; algorithmic bias

Uvod

Umetna inteligenca (UI) ni več znanstvena fantastika, temveč del našega vsakdana. Spreminja način, kako delamo, se učimo in razmišljamo. Ta preobrazba je v zadnjih letih še posebej očitna z razvojem generativne umetne inteligence (Gen-UI). Orodja, kot so veliki jezikovni modeli (angl. *Large Language Models – LLM*), na primer ChatGPT, Llama, Gemini, DeepSeek in druga, ter orodja za ustvarjanje slik, kot je DALL-E, so hitro prodrla v šole, univerze in učilnice. Pojavila se je spontano, brez predhodnega vrednotenja ali načrtovanja, kar prinaša velike priložnosti za individualizacijo učenja, hkrati pa tudi različne izzive, ki so povezana z rezultati učenja, digitalnim razkorakom ipd. Empirične raziskave nedvomno kažejo, da lahko Gen-UI, ko je premišljeno vključena, znatno izboljša učne dosežke in zavzetost, zlasti pri tistih učencih, ki so bili doslej v izobraževalnem sistemu pogosto prezrti (Hmoud idr., 2024; Jauhiainen in Guerra, 2023).

Vendar pa ta najnovejša tehnološka revolucija ni brez pasti. Z njo prihajajo resni izzivi, ki jih podrobneje omenjamo tudi v posameznih poglavjih te knjige: od t.i. »halucinacij« (generiranja napačnih informacij) in algoritmične pristranskosti, ki lahko okrepi obstoječe družbene neenakosti, do »kognitivnega dolga«, kjer prekomerna odvisnost od UI oslabi sposobnost kritičnega mišljenja in globljega razumevanja. Kljub vsemu pa ni več vprašanje, ali bomo UI uporabljali v izobraževanju, temveč kako. Ali jo bomo dojemali kot orodje za nadomestitev učiteljev, kot nevarnost, ki jo je treba prepovedati, ali kot zmogljivega pomočnika, ki bo človeško znanje šele dvignil na višjo raven?

Kaj sploh je umetna inteligenca?

Umetna inteligenca (UI) je ena najvplivnejših tehnologij našega časa, vendar jo pogosto napačno razumemo. Ne gre za eno samo tehnologijo, temveč za široko področje računalništva, ki se ukvarja s sistemi, sposobnimi izvajanja nalog, ki običajno zahtevajo človeško inteligenco. Sodobne definicije, denimo tista

organizacije OECD (2026), UI opredeljujejo kot sistem, ki na podlagi strojnih ali človeških ciljev sprejema sklepe o tem, kako vplivati na okolje s priporočili, napovedmi ali odločitvami.

Ključna značilnost sodobne UI je prehod od *programiranja pravil* k *učenju iz podatkov*. Klasični računalniški programi delujejo po vnaprej zapisanih korakih, medtem ko sistemi UI (predvsem strojno učenje) samostojno izluščijo vzorce iz ogromne količine primerov. To jim omogoča, da se prilagajajo novim situacijam brez neposrednega človekovega posega.

Zmožnosti teh sistemov danes segajo od preprostega prepoznavanja vzorcev do kompleksne obdelave naravnega jezika. Veliki jezikovni modeli, kot je ChatGPT, lahko generirajo besedila, prevajajo in celo pišejo programsko kodo. Kljub temu je treba razumeti pomembno razliko: čeprav ti sistemi »simulirajo« razumevanje, v resnici le statistično obdelujejo podatke (Sejnowski, 2023).

Zato v strokovni literaturi strogo ločimo med dvema vrstama inteligence:

- **Ozka umetna inteligenca (artificial narrow intelligence; ANI):** To so vsi današnji sistemi. So izjemno zmogljivi, a specializirani le za specifične naloge (npr. igranje šaha, vožnja avtomobila ali pisanje eseja).
- **Splošna umetna inteligenca (artificial general intelligence; AGI):** To je hipotetična inteligenca, ki bi bila sposobna reševati katerokoli intelektualno nalogo enako dobro kot človek. Takšna tehnologija za zdaj še ne obstaja (Jones, 2024).

Zgodovinski razvoj

Zgodovina umetne inteligence ni premočrtna pot napredka, temveč dinamično zaporedje obdobij velikega navdušenja, ki so jim sledila obdobja razočaranja in stagnacije. Da bi razumeli današnjo revolucijo, se moramo vrniti k samim temeljem računalništva.

Od filozofskih vprašanj do rojstva discipline (1950–1956)

Še preden je področje dobilo ime, je Alan Turing leta 1950 v prelomnem članku *Computing Machinery and Intelligence* zastavil ključno vprašanje: »Ali lahko stroji mislijo?« Predlagal je slavni »Turingov test« kot merilo, s katerim bi ocenili sposobnost stroja za inteligentno vedenje, nerazločljivo od človeškega (Jones, 2024). Formalno rojstvo discipline pa sega v leto 1956, ko je John McCarthy na konferenci v Dartmouthu zbral vodilne raziskovalce in prvič uporabil izraz **umetna inteligenca**. V tem obdobju je prevladoval optimizem; raziskovalci so verjeli, da bodo z uporabo logičnih pravil in simbolov kmalu simulirali vse vidike človeškega mišljenja.

Ovire pri razvoju (1970–1980)

Prvotni pristopi, ki so temeljili na strogo določenih pravilih (t. i. simbolna UI), so v nadzorovanih laboratorijskih okoljih delovali obetavno, a so hitro naleteli na ovire, ko so jih soočili s kompleksnostjo resničnega sveta. Računalniki tistega časa niso imeli zadostne računske moči, programiranje vseh možnih pravil za vsako situacijo pa se je izkazalo za nemogoče. Neizpolnjene obljube so v 70. letih vodile do drastičnega zmanjšanja financiranja raziskav – obdobja, ki ga danes poznamo kot »zima umetne inteligence« (Russell in Norvig, 2021; Jones, 2024).

Nastanek strojnega učenja (1980–2010)

V 80. in 90. letih je prišlo do ključnega miselnega preskoka. Namesto da bi računalnikom *programirali* pravila, kako naj rešijo problem, so raziskovalci začeli razvijati sisteme, ki se *učijo* iz podatkov. To je bil vzpon strojnega učenja (ang. *machine learning*). Uporaba statističnih metod in ponovno odkritje nevronskih mrež sta omogočila napredek, vendar je bil ta še vedno omejen s količino dostopnih podatkov.

Razvoj globokega učenja in transformatorjev (2012–danes)

Pravi preboj, ki je oblikoval sedanji trenutek, se je zgodil leta 2012. Model AlexNet je z uporabo globokega učenja (ang. *deep learning*) dosegel izjemno natančnost pri prepoznavanju slik. Osnova za ta preboj je bila nova arhitektura grafičnih kartic, ki je omogočala vzporedno procesiranje podatkov, kar s pridom izkoriščajo nevronske mreže. To je bil dokaz, da lahko UI z dovolj podatki in računske moči preseže človeške sposobnosti pri specifičnih nalogah (Sejnowski, 2023).

Zadnje veliko poglavje pa se piše od leta 2017 z razvojem arhitekture **transformatorjev** (ang. *transformers*). Ta tehnologija je omogočila, da se modeli ne učijo le posameznih besed, temveč »razumejo« kontekst in odnose med besedami v dolgih besedilih. Prav transformatorji so temelj velikih jezikovnih modelov (LLM), kot sta GPT-5 in Claude, ki danes generirajo smiselna besedila, pišejo kodo in ustvarjajo umetnost, kar je bilo še pred desetletjem nepredstavljivo (Sejnowski, 2023).

Zmogljivosti in omejitve današnje UI

Da bi umetno inteligenco v šoli uporabljali učinkovito, moramo natančno razumeti, kaj ta tehnologija zmore in – še pomembneje – kje so njene meje. Pogosto namreč nasedamo iluziji, da imamo opravka z inteligentnim sogovornikom, medtem ko gre v resnici za statistični stroj.

V strokovni literaturi poznamo temeljno delitev na ozko (ANI) in splošno (AGI) umetno inteligenco. Vsi sistemi, ki jih uporabljamo danes – od algoritmov na družbenih omrežjih do naprednih modelov, kot je GPT-5 –, sodijo v kategorijo ozke umetne inteligence. To pomeni, da so izjemno uspešni pri opravljanju specifičnih nalog, za katere so bili trenirani, vendar nimajo širšega razumevanja sveta, zavesti ali zmožnosti samostojnega prenašanja znanja na povsem nova, nepovezana področja (Jones, 2024). Čeprav se zdi, da sodobni modeli obvladajo »vse«, od pisanja pesmi do programiranja, gre tehnično gledano le za »modele za splošne namene« (ang. *general purpose AI*), ki še vedno ne dosegajo prave človeške kognitivne prožnosti.

Največja omejitev današnje UI je, da sveta ne razume vzročno-posledično, temveč korelativno. Ko model napiše esej o francoski revoluciji, ne »ve«, kaj je revolucija, in ne razume trpljenja ali političnih sprememb. Zanaša se izključno na prepoznavanje vzorcev in statističnih povezav med besedami v svoji bazi podatkov (Sejnowski, 2023). To pojasnjuje, zakaj lahko isti model reši kompleksen matematični problem (ker je videl podobne vzorce), a nato pade pri preprostem logičnem vprašanju, ki zahteva »zdravo kmečko pamet«, kakršne nimamo zapisane v učbenikih.

Za učitelje je ključno razumevanje koncepta t. i. nazobčane meje (ang. *jagged frontier*), ki ga opisuje Mollick (2024) in po katerem sposobnosti UI niso enakomerne. Pri nekaterih težkih nalogah (npr. generiranje idej za učno uro, pisanje kode) je UI nadpovprečno uspešna, pri drugih, ki se ljudem zdijo trivialne (npr. štetje števila besed v stavku ali preverjanje točnosti citatov), pa pogosto odpove. Ta nepredvidljivost pomeni, da tehnologije ne moremo uporabljati na pamet, temveč moramo vsak njen izdelek preveriti.

Kljub omejitvam so zmogljivosti sodobnih sistemov impresivne. Ključna novost je večmodalnost – sposobnost sočasnega procesiranja in generiranja besedil, slik, zvoka in celo videa. To odpira vrata do prilagojenih učnih gradiv, kjer lahko učitelj v nekaj sekundah ustvari vizualno razlago kompleksnega koncepta ali prilagodi besedilo bralnim sposobnostim posameznega učenca.

Kaj pa generativna umetna inteligenca?

Če je tradicionalna umetna inteligenca zadnjih desetletij veljala za zmogljivega analitika, je generativna umetna inteligenca (Gen-UI) stopila na oder kot ustvarjalec. Gre za podskupino umetne inteligence, ki pomeni enega največjih tehnoloških preskokov v zgodovini računalništva. Medtem ko so prejšnje generacije algoritmov podatke predvsem analizirale, razvrščale ali filtrirale, ima generativna UI sposobnost ustvarjanja povsem novih, izvirnih vsebin.

Da bi razumeli to novost, moramo potegniti ločnico med *diskriminativno* (analitično) in *generativno* UI.

- Analitično UI poznamo že dolgo: sistem, ki se nauči razlikovati med nezaželeno in pomembno e-pošto, ali algoritem, ki na rentgenski sliki prepozna zlom. Njen cilj je klasifikacija obstoječih podatkov.
- Generativna UI pa gre korak dlje. Ne le, da prepozna vzorce v podatkih, temveč te vzorce uporabi za gradnjo novih primerov, ki v resničnem svetu prej niso obstajali. Na podlagi naučenih pravil statistične verjetnosti lahko ustvari nov esej, unikatno sliko, simfonijo ali delujočo programsko kodo, na las podobno tisti, ki bi jo napisal človek.

Kako to deluje? V jedru generativne UI so globoke nevronske mreže, ki so se urile na nepredstavljivih količinah podatkov (npr. celotnem javno dostopnem svetovnem spletu). Med tem procesom model ne »memorizira« podatkov kot enciklopedija, temveč gradi kompleksne notranje reprezentacije o tem, kako je svet sestavljen, na primer katere besede pogosto stojijo skupaj, kako se svetloba odbija na fotografiji ali kako so strukturirani stavki v določenem jeziku.

Ko modelu podamo navodilo (ang. *prompt*), ta ne išče obstoječega odgovora v bazi podatkov (kot počne iskalnik Google), temveč besedo za besedo ali piksel za piksel generira nov odgovor. To pojasnjuje, zakaj lahko ChatGPT na isto vprašanje vsakič odgovori malce drugače. Gre za verjetnostni proces ustvarjanja, ne za deterministični proces priklica informacij (Sejnowski, 2023).

Za šolski prostor je ta prehod monumentalen. Tehnologija ni več le orodje za dostop do informacij, temveč postaja orodje za *produkcijo* informacij. Kot poudarja UNESCO (2023), to temeljito spreminja vlogo učenca in učitelja: fokus se premika od pomnjenja in iskanja virov k sposobnosti kritičnega vrednotenja in usmerjanja stroja pri ustvarjanju vsebin.

Temeljne značilnosti generativne umetne inteligence

Srce generativne revolucije so t. i. temeljni modeli (ang. *Foundation Models*). Gre za sisteme, ki so trenirani na tako obsežnih podatkih (velik del svetovnega spleta, digitalizirane knjige, kode, slike), da jih je mogoče prilagoditi za opravljanje neštetih različnih nalog. Za razliko od prejšnjih sistemov, ki so bili »specialisti« (npr. samo za prevajanje), so generativni modeli »generalisti« (Stanford HAI, 2024). Njihove ključne značilnosti, ki spreminjajo izobraževalno okolje, so: interaktivnost in naravni jezik, večmodalnost (ang. *multimodality*), prilagodljivost in »učenje v trenutku«.

Največja sprememba, ki jo je prinesel ChatGPT, ni bila le pametnejša tehnologija, ampak vmesnik. Generativna UI omogoča komunikacijo v naravnem, pogovornem jeziku. Učitelju ali učencu ni treba znati programirati; modelu

preprosto poda navodilo ali »poziv« (ang. *prompt*). Model je sposoben ohranjati kontekst pogovora, kar omogoča postopno izboljševanje rezultatov s pomočjo dialoga – proces, ki ga imenujemo iterativno sodelovanje.

Medtem ko so bili prvi modeli omejeni le na besedilo, je sodobni standard, že omenjena, večmodalnost. To pomeni, da lahko isti model hkrati obdeluje in generira različne vrste medijev. Uporabnik lahko modelu naloži fotografijo matematične enačbe in prosi za rešitev ali pa na podlagi skice ustvari delujočo spletno stran. Kot ugotavlja poročilo AI Index (Stanford HAI, 2024), prav ta prehod omogoča, da UI postane resnično vsestranski asistent pri ustvarjanju gradiva.

Generativni modeli so sposobni t. i. učenja z nekaj primeri (ang. *few-shot learning*). Če modelu pokažemo tri primere dobrega povzetka, bo četrtič razumel naš slog in ga posnemal, ne da bi ga morali za to posebej programirati. To učiteljem omogoča, da orodja hitro prilagodijo svojim specifičnim potrebam (Sejnowski, 2023).

Med generativnimi modeli Sejnowski (2023) omenja:

- ChatGPT (generiranje besedil in kode),
- DALL-E (ustvarjanje slik iz besedilnih opisov),
- Stable Diffusion in Midjourney (ustvarjanje umetniških in fotorealističnih slik),
- Codex in Copilot (generiranje programske kode) (Sejnowski, 2023).

Čeprav so ta orodja impresivna, je nujno razumeti, da njihova »ustvarjalnost« temelji na statističnem kombiniranju obstoječega znanja, kar nas pripelje do vprašanja, kako ti modeli v resnici »razmišljajo«.

Kako »razmišlja« veliki jezikovni model?

Ko se pogovarjamo s ChatGPT ali podobnim orodjem, imamo močan občutek, da je na drugi strani inteligentno bitje, ki razume naše vprašanje, poišče odgovor v svojem spominu in nam ga posreduje. Vendar je ta antropomorfizacija – pripisovanje človeških lastnosti stroju – nevarna. Da bi razumeli omejitve orodja, moramo pogledati pod pokrov motorja.

Veliki jezikovni model (LLM) ne deluje kot baza podatkov ali spletni iskalnik. Iskalnik (npr. Google) hrani indeks spletnih strani in nam vrne tisto, ki se ujema z našim iskanjem. Jezikovni model pa ne »hrani« besedil. Namesto tega hrani milijarde matematičnih povezav (parametrov) med delčki besed, ki so se vzpostavile med ustvarjanjem modela.

V svojem bistvu je LLM stroj za napovedovanje naslednjega žetona (ang. *next-token prediction*). Besedilo razbije na manjše enote, imenovane žetoni (ang. *tokens*) – to so lahko cele besede, zlogi ali celo posamezne črke. Njegova edina

naloga je, da na podlagi vseh prejšnjih žetonov v pogovoru izračuna, kateri žeton bo z največjo verjetnostjo sledil (Wolfram, 2023). Proces si lahko predstavljamo kot izjemno napredno funkcijo samodokončanja besedila na pametnem telefonu, le da model ne gleda le zadnje besede, temveč upošteva celoten kontekst pogovora, slog, ton in celo logično strukturo, ki jo je zaznal v učnih podatkih.

Pomembna termina sta tudi koncepta »verjetnosti« in »temperature«. Važno je razumeti, da model ne izbere vedno najverjetnejše besede. Če bi vedno izbral le statistično najpogostejšo naslednjo besedo, bi bilo besedilo suhoparno, ponavljajoče se in nenaravno. Zato modeli uporabljajo parameter, imenovan temperatura, ki v proces vnese element naključnosti. To modelu omogoča, da je »kreativen« in tvori stavke, ki jih prej še nikoli ni videl. Ravno ta mehanizem pa je dvorezni meč (Wolfram, 2023) in vodi do pojava, ki ga poljudno imenujemo »halucinacije«, strokovno pa konfabulacije. Ker model ne operira s konceptom resnice, temveč le z verjetnostjo, lahko ustvari zaporedje besed, ki je slovnično in slogovno brezhibno, a vsebinsko popolnoma izmišljeno. Za model razlika med dejstvom in prepričljivo lažjo ne obstaja – oboje je le statistično verjeten niz besed.

Ko model vprašamo nekaj, o čemer nima veliko podatkov, bo kljub temu poskušal sestaviti stavek, ki zveni prepričljivo. Ker so v njegovem učnem nizu besede, kot so »znanstvena študija«, »leta 2023« in »objavljeno v reviji Nature«, pogosto povezane, jih bo morda združil hin si izmislil povsem neobstoječo raziskavo. Za model to ni laž, saj nima namena zavajati; to je le statistično verjetno nadaljevanje stavka (Sejnowski, 2023).

Za šolsko prakso to pomeni pomembno spoznanje: UI je odlična pri nalogah, ki zahtevajo ustvarjalnost, preoblikovanje in iskanje idej, vendar ni zanesljiva pri nalogah, kjer je potrebna faktografska natančnost. Model ne preverja dejstev, temveč zgolj ustvarja besedilo, ki zveni prepričljivo.

Je umetna inteligenca nepristranska?

Ena najnevarnejših zmot o umetni inteligenci je prepričanje, da so računalniški algoritmi objektivni, ker temeljijo na matematiki in logiki. Pogosto slišimo: »To je izračunal računalnik, torej je res.« V resnici pa umetna inteligenca ne deluje v vakuumu. Je produkt človeških odločitev in podatkov, ki so vse prej kot nevtralni. UI ne presoja sveta skozi sito resnice, temveč deluje kot zrcalo družbe, ki jo je ustvarila – z vsemi njenimi vrlinami in, žal, tudi z vsemi njenimi predsodki.

Temeljni vzrok pristranskosti tiči v podatkih. Veliki jezikovni modeli se učijo na milijardah besedil, pobranih s spleta. Svetovni splet pa ni reprezentativen odraz celotnega človeštva; je odraz tistih, ki so v zgodovini imeli moč in dostop do tehnologije.

Kot opozarja Kate Crawford (2021), učni nabori pogosto vsebujejo pretežno zahodne, anglocentrične perspektive (t. i. WEIRD populacija – *Western, Educated, Industrialized, Rich, Democratic*). Posledično modeli pogosto reproducirajo stereotipe, npr.:

- **Spolni stereotipi:** Če model prosimo za zgodbo o zdravniku in medicinski sestri, bo zdravnika pogosto poimenoval z moškim imenom, sestro pa z ženskim, saj se ti vzorci statistično pogosteje pojavljajo v starih besedilih.
- **Kulturna dominacija:** Model bo morda odlično poznal ameriško zgodovino, dogodke iz slovenske ali afriške zgodovine pa bo ignoriral ali interpretiral skozi prizmo ameriške kulture.

Druga raven pristranskosti vstopi v proces med t. i. »vzpodbujevalnim učenjem« (ang. *RLHF – Reinforcement Learning from Human Feedback*). Da bi bili modeli varni in vljudni, njihove odgovore ocenjujejo ljudje. Ti ocenjevalci imajo lastne kulturne vrednote, ki se prenesejo na model. Kar je v eni kulturi vljuden odgovor, je lahko v drugi žaljiv. UI torej nima lastne morale; ima »vgrajeno moralo« specifične skupine ljudi, ki je model razvijala in trenirala (UNESCO, 2023).

Za učitelje in učence to pomeni, da rezultatov UI ne smemo nikoli sprejemati kot nevtralnih dejstev. Umetna inteligenca lahko nenamerno okrepi obstoječe družbene neenakosti ali izključi manjšinska stališča. Zato ključna kompetenca prihodnosti ni le uporaba UI, temveč kritično vrednotenje njenih rezultatov – prepoznavanje, čigav glas je v odgovoru zastopan in čigav utišan.

Nevarnost kognitivnega dolga

To je verjetno najpomembnejše opozorilo za učitelje, saj se neposredno dotika našega največjega strahu: »Bodo otroci nehali razmišljati?«

S prihodom orodij, ki znajo pisati, programirati in reševati naloge, se v izobraževanju odpira ključno vprašanje: Kaj se zgodi z našimi možgani, če težke miselne naloge prepustimo stroju? Pojav, ki ga opisujemo kot »kognitivni dolg« ali »kognitivno razbremenjevanje« (ang. *cognitive offloading*), ni nov. Ko smo začeli uporabljati kalkulatorje, smo deloma izgubili spretnost hitrega računanja na pamet; ko smo sprejeli GPS navigacijo, smo oslabili svojo sposobnost orientacije v prostoru.

Vendar generativna umetna inteligenca prinaša tveganje na povsem novi ravni. Ne nadomešča le rutinskih operacij, temveč posega v procese, ki so temelj učenja: sintezo informacij, strukturiranje argumentov in kritično razmišljanje.

Učenje ni zgolj zbiranje informacij, temveč proces gradnje nevronskih povezav, ki nastajajo z miselnim naporom. Ko se učeči trudi s pisanjem prvega odstavka eseja ali iskanjem napake v kodi, v njegovih možganih poteka globoko učenje.

Če celotno nalogo prepustimo ChatGPT, se odrečemo miselni vadbi. Ethan Mollick (2024) to imenuje »past avtomatizacije«: ko UI uporabljamo kot nadomestek za razmišljanje in ne kot orodje za njegovo razširitev, tvegamo izgubo kognitivnih sposobnosti. Učeči se sicer hitro pride do pravilnega odgovora, a ne razvije procesnega znanja – ne ve, kako bi do rešitve prišel sam. Dolgoročno to pomeni več kot le manko posameznih veščin: kot poudarjata Hiebert in Grouws (2007), učeči se, ki se ne soočajo z izzivi, ne razvijejo vztrajnosti pri reševanju težav in tolerance do negotovosti, ki sta pomembna za samostojno učenje. Brez procesnega razumevanja tudi težko prenesejo naučeno v nove kontekste – znanje ostane vezano na konkretno situacijo, namesto da bi ga lahko uporabili tudi v novih situacijah.

S tem je povezana nevarnost iluzije kompetenc. Ker je odgovor UI jasen in berljiv, ima uporabnik lažen občutek, da snov razume. Prebiranje odlično napisanega eseja, ki ga je ustvaril stroj, ne pomeni, da bi ga učenec znal napisati sam. UNESCO (2023) opozarja, da mora izobraževalni sistem zagotoviti, da tehnologija ne oslabi naše sposobnosti samostojnega delovanja in sprejemanja odločitev. Cilj torej ni prepoved tehnologije, saj bi s tem učence prikrajšali za pomembne digitalne veščine. Rešitev leži v spremembi načina uporabe. UI ne sme začeti in končati miselnega procesa.

Zaključek

Pregled, ki smo ga opravili v tem poglavju in je segal od prvih logičnih modelov do današnjih nevronskih mrež, je pokazal, da umetna inteligenca ni več stvar oddaljene prihodnosti, temveč realnost, ki korenito preoblikuje naše delo in učenje. Kot smo spoznali, generativna umetna inteligenca prinaša izjemno ustvarjalnost in prilagodljivost, hkrati pa odpira težka vprašanja o pristranskosti, halucinacijah in kognitivni prikrajšanosti.

Zelo zagovarjamo stališče, da učitelj ostaja srce učnega procesa – kot mentor, moderator in vodja, ki ga stroj, ne glede na svoje zmogljivosti, ne more nadomestiti.

Literatura

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Hiebert, J., & Grouws, D. A. (2007). The Effects of Classroom Mathematics Teaching on Students' Learning. V F. K. Lester Jr. (Ur.), *Second Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics* (str. 371–404). Information Age Publishing.

Hmoud, M., Swaity, H., Hamad, N., Karram, O. in Daher, W. (2024). Higher Education Students' Task Motivation in the Generative Artificial Intelligence Context: The Case of ChatGPT. *Information*, *15*(1), 33. https://doi.org/10.3390/info15010033

Jauhiainen, J. S. in Guerra, A. G. (2023). Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability*, *15*(18), 14025. https://doi.org/10.3390/su151814025

Jones, B. (2024). *AI Literacy Fundamentals.* Data Literacy Press.

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Portfolio/Penguin.

Russell, S. in Norvig, P. (2021). *Artificial Intelligence: A Modern Approach (4. izd.).* Pearson.

OECD (2026). OECD *Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*. OECD Publishing, Paris, https://doi.org/10.1787/062a7394-en

Sejnowski, T. J. (2023). *ChatGPT and the Future of AI: The Deep Language Revolution*. MIT Press.

Stanford HAI (2024). *The AI Index 2024 Annual Report.* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. Dostopno na: https://hai.stanford.edu/ai-index/2024-ai-index-report

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *49*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

UNESCO (2023). *Guidance for generative AI in education and research.* UNESCO. https://doi.org/10.54675/EWZM9535

Wolfram, S. (2023). *What is ChatGPT doing ... and why does it work?.* Stephen Wolfram Writings. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/