■ *Pregledni znanstveni članek*

Umut Arıöz, Grega Močnik

Agenti umetne inteligence v zdravstvu

Povzetek. Agenti umetne inteligence predstavljajo nadaljnji razvoj umetne inteligence s premikom od sistemov, ki se odzivajo izključno na uporabniške pozive, k sistemom, ki so sposobni samostojnega načrtovanja, izvajanja in prilagajanja delovanja glede na zastavljene cilje. Temeljijo na velikih jezikovnih modelih, ki delujejo kot osrednje spoznavno jedro sistema, nadgrajeno z mehanizmi za uporabo orodij in različnimi oblikami spomina. Takšna zasnova omogoča večstopenjsko razmišljanje, ponavljalno vrednotenje lastnih dejanj ter samostojno izvajanje kompleksnih nalog. Agenti delujejo v neprekinjenem krogu zaznavanja, načrtovanja, ukrepanja in razmisleka, kar jih jasno loči od običajnih jezikovnih modelov. V praksi se agenti že uveljavljajo na področjih raziskovanja, razvoja programske opreme in poslovnih procesov, kjer prevzemajo zbiranje informacij, analizo podatkov in avtomatizacijo odločanja. Poseben pomen imajo v zdravstvu, kjer prispevajo k razbremenitvi zdravnikov z avtomatizacijo dokumentacije, podporo kliničnemu odločanju, analizo medicinskih podatkov, izvajanjem triaže in optimizacijo organizacije dela. Hkrati z večjo stopnjo samostojnosti narašča pomen etičnih in pravnih vprašanj, povezanih z odgovornostjo, varnostjo in nadzorom nad delovanjem agentov. Njihova zanesljiva uporaba zahteva premišljeno regulacijo, transparentne mehanizme delovanja in jasno opredeljen človeški nadzor.

Ključne besede: agenti umetne inteligence, klinično odločanje, zdravstvena oskrba, etika, avtomatizacija dokumentacije.

Artificial Intelligence Agents in Health Care

Abstract. Artificial intelligence agents represent a further stage in the development of artificial intelligence, moving from systems that respond solely to user prompts to systems capable of autonomous planning, action and adaptation to predefined goals. They are based on large language models that serve as the core cognitive component, extended with mechanisms for tool use and different forms of memory. This architecture enables multi-step reasoning, iterative self-evaluation and autonomous execution of complex tasks. The operation of agents follows a continuous cycle of perception, planning, action and reflection, which clearly distinguishes them from conventional language models. In practice, agents are increasingly applied in research, software development and business processes, where they perform information gathering, data analysis and decision automation. In health care, they contribute to reducing physicians' workload through automated documentation, clinical decision support, medical data analysis, patient triage and optimisation of organisational workflows. With increasing autonomy, ethical and legal issues related to responsibility, safety and oversight become more prominent. Reliable deployment therefore requires careful regulation, transparency and clearly defined human supervision.

Key words: artificial intelligence agents, clinical decision-making, healthcare delivery, ethics, documentation automation.

■ **Infor Med Slov** 2025; 30(1-2): 28-35

Institucija avtorjev / Authors' institution: Laboratorij za digitalno procesiranje signalov, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, Maribor (UA, GM).

Kontaktna oseba / Contact person: dr. Umut Arıöz, Laboratorij za digitalno procesiranje signalov, Fakulteta za elektrotehniko, računalništvo in informatiko,, Koroška cesta 46, 2000 Maribor, Slovenija. E-pošta / E-mail: umut.arioz@um.si

Prispelo / Received: 9. 12. 2025. Sprejeto / Accepted: 6. 1. 2026.

Uvod

Agenti umetne inteligence (v nadaljevanju: agenti UI) so programski sistemi, zasnovani za samostojno doseganje ciljev v dinamičnem okolju z uporabo zaznavanja, načrtovanja in izvajanja dejanj. Za razliko od tradicionalnih sistemov umetne inteligence, ki delujejo pretežno reaktivno, agenti UI aktivno usmerjajo svoje delovanje glede na zastavljene cilje in povratne informacije iz okolja.

V praksi si lahko agenta UI predstavljamo kot digitalni sistem, ki mu je dodeljena naloga, na primer priprava celovitega poročila o obremenitvah ambulante na podlagi razpoložljivih podatkov. Agent samostojno zbere ustrezne informacije iz informacijskih sistemov, jih analizira, ovrednoti rezultate in pripravi strukturirano poročilo, ki ga lahko uporabnik pregleda in potrdi. Tak pristop presega enkratno ustvarjanje odgovorov in predstavlja osnovo za samostojno podporo zahtevnim delovnim procesom.

Od jezikovnih modelov do agentov UI

Razumevanje agentov UI zahteva jasno razlikovanje med velikimi jezikovnimi modeli (angl. *Large Language Models – LLM*) in agentskimi sistemi. V zadnjem desetletju so razvoj in uporabo umetne inteligence v veliki meri zaznamovali veliki jezikovni modeli, med katerimi sta najbolj prepoznavna modela GPT in BERT. Ti modeli so zasnovani za obdelavo in tvorjenje besedilnih vsebin na podlagi statističnih vzorcev v obsežnih učnih podatkovnih zbirkah ter dosegajo visoko stopnjo uspešnosti pri razumevanju naravnega jezika, povzetkih besedil, pisanju programske kode in sklepanju iz daljših strokovnih vsebin.[1,2] Njihovo delovanje temelji na verjetnostni napovedi zaporedja besed, kar omogoča tvorjenje koherentnih in semantično ustreznih besedilnih izhodov.[2]

Kljub naprednim zmožnostim so veliki jezikovni modeli po svoji arhitekturi in namenu pretežno reaktivni sistemi. Delujejo v okviru posamezne interakcije, kjer na podan vhod ustvarijo enkraten izhod, nato pa se njihovo delovanje zaključi. Takšni modeli nimajo notranjega mehanizma za dolgoročno načrtovanje, vztrajanje pri cilju ali samostojno prilagajanje strategije na podlagi zaporednih povratnih informacij. Posledično so omejeni na vlogo naprednega orodja za obdelavo jezika, brez neposredne sposobnosti samostojnega delovanja v kompleksnih in dinamičnih okoljih.[3]

Agenti UI predstavljajo arhitekturno in funkcionalno nadgradnjo tega pristopa. Pri agentskih sistemih veliki jezikovni model ne deluje kot samostojen končni sistem, temveč kot osrednje miselno oziroma spoznavno jedro širšega programskega okvira. Ta okvir vključuje dodatne zmogljivosti, kot so dostop do zunanjih orodij, uporaba programskih vmesnikov (angl. *Application Programming Interface – API*), izvajanje kode ter vzdrževanje različnih oblik spomina. S tem agent UI preseže omejitve zgolj jezikovne obdelave in postane sistem, sposoben zaznavanja okolja, zaporednega odločanja in izvajanja dejanj v digitalnem ali realnem okolju.[4]

Bistvena razlika med velikim jezikovnim modelom in agentom UI je v stopnji proaktivnosti in vztrajnosti delovanja. Agentu UI je dodeljen cilj, ki ga ne doseže z enim samim odzivom, temveč skozi zaporedje premišljenih korakov. Pri tem uporablja jezikovni model za razumevanje naloge, razčlenitev cilja na podnaloge, vrednotenje vmesnih rezultatov ter sprotno prilagajanje nadaljnjega delovanja. Takšna zasnova omogoča večstopenjsko razmišljanje, reševanje kompleksnih problemov in sprejemanje odločitev na podlagi iterativnega vrednotenja lastnih dejanj in njihovih posledic.[5]

Prehod od običajnih velikih jezikovnih modelov k agentskim sistemom pomeni spremembo v vlogi umetne inteligence. Namesto pasivnega orodja za obdelavo informacij agenti UI delujejo kot samostojni programski sistemi, ki so sposobni izvajati celovite naloge, usklajevati več virov informacij in se prilagajati spremenljivim okoliščinam. Ta razvoj predstavlja pomembno osnovo za uporabo umetne inteligence v okoljih, kjer so potrebni dolgoročnejše načrtovanje, zanesljivo izvajanje postopkov in nadzor nad kompleksnimi procesi.[5]

Opredelitev samostojnosti

Samostojnost agentov UI temelji na njihovem delovanju v neprekinjenem povratnem krogu, ki omogoča samostojno doseganje zastavljenih ciljev v dinamičnem okolju.[6,7] Gre za obliko samostojne programske opreme, zasnovane tako, da svoje delovanje organizira skozi zaporedje štirih temeljnih procesov:

- zaznavanje,
- načrtovanje,
- ukrepanje in
- razmislek.

V fazi zaznavanja agent UI pridobi informacije o zastavljenem cilju ter o stanju okolja, v katerem deluje.

To lahko vključuje zbiranje podatkov iz informacijskih sistemov, analiziranje vhodnih signalov ali pridobivanje posodobljenih informacij iz zunanjih virov. Na tej osnovi agent oblikuje trenutno predstavo o problemu, ki ga mora obravnavati.

Sledi faza načrtovanja, v kateri agent UI razčleni kompleksno nalogo na zaporedje manjših, izvedljivih korakov. Pri tem veliki jezikovni model deluje kot osrednje spoznavno jedro sistema in podpira kognitivne funkcije, kot so razumevanje cilja, interpretacija navodil, oblikovanje kontekstualnega okvira ter zasnova zaporednih dejanj, ki vodijo k doseganju cilja.[7,8] Načrtovanje tako presega enkratno odzivanje in vključuje strukturirano odločanje v več korakih.

V fazi ukrepanja agent UI izvede načrtovane korake z uporabo razpoložljivih orodij. To lahko vključuje uporabo programskih vmesnikov, izvajanje kode, dostop do podatkovnih zbirk ali interakcijo z drugimi digitalnimi sistemi. Sposobnost uporabe zunanjih orodij predstavlja ključen pogoj za dejansko avtonomijo, saj agentu omogoča izvajanje dejanj, ki presegajo zgolj obdelavo ali tvorjenje besedila.[7,9]

Faza razmisleka omogoča vrednotenje izvedenih dejanj in njihovih rezultatov. Agent UI analizira, ali so bila dejanja uspešna glede na zastavljeni cilj, ter na tej osnovi prilagodi nadaljnje načrtovanje in ukrepanje. Ta mehanizem samoocenjevanja je temelj za iterativno izboljševanje delovanja in omogoča postopno prilagajanje strategije, če se okoliščine spremenijo ali agent zaznana napako.[9]

Takšen neprekinjen delovni krog omogoča agentom UI vztrajno delovanje, sprotno popravljanje lastnih odločitev in prilagajanje novim informacijam. Pomemben sestavni del samostojnosti je tudi uporaba spomina, ki agentu omogoča ohranjanje koherence in učenje skozi čas. Kratkoročni oziroma kontekstualni spomin podpira razumevanje trenutne naloge in zaporedja interakcij, epizodni spomin shranjuje zapise preteklih dejanj in povratnih informacij, dolgoročni oziroma zunanji spomin pa omogoča dostop do strukturiranega znanja iz podatkovnih baz ali drugih trajnih virov.[9]

Neprekinjena povratna zanka zaznavanja, načrtovanja, ukrepanja in razmisleka predstavlja osrednji mehanizem avtonomnega delovanja agentov UI. Ta zasnova omogoča obvladovanje kompleksnih nalog, ki zahtevajo zaporedno odločanje, prilagajanje in dolgoročnejše sledenje ciljem, ter predstavlja osnovo za njihovo uporabo v zahtevnih aplikacijskih okoljih.[7]

Nove oblike uporabe

Opisani način delovanja agentov umetne inteligence omogoča obravnavo izjemno zapletenih nalog, kar odpira nove možnosti njihove uporabe. Takšne agentske sisteme lahko glede na osnovno usmeritev delovanja razdelimo na agente, usmerjene v sklepanje in razvrščanje, agente, usmerjene v simulacijo, ter interaktivne pogovorne agente. Agenti UI se lahko vključujejo v ključno infrastrukturo in kompleksne sisteme. Uporabljajo se na primer v simulacijah avtonomnega krmiljenja vesoljskih plovil ali kot podpora operacijam v zahtevnih raziskovalnih okoljih, kot so pospeševalniki delcev, kjer povezujejo orodja za pridobivanje znanja in upravljanje strojev.[7,9]

Tovrstna uporaba predstavlja prehod med splošnimi agentskimi sistemi in specializiranimi rešitvami, ki se vse pogosteje uporabljajo tudi v zdravstvenih informacijskih okoljih.

Praktična uporaba

Agenti UI se vse pogosteje uporabljajo v okoljih, kjer je potrebna obdelava velikih količin podatkov, sprotno prilagajanje na nove informacije in zaporedno odločanje. V nasprotju s tradicionalnimi sistemi avtomatizacije, ki temeljijo na vnaprej določenih pravilih, agenti UI delujejo bolj prilagodljivo, saj se lahko odzivajo na spremembe v vhodnih podatkih in ciljih brez neposrednega posredovanja uporabnika. Njihova uporaba danes ni več omejena na raziskovalna ali eksperimentalna okolja, temveč postajajo del vsakdanjih praks na različnih strokovnih področjih.[10,11]

Pomemben premik je opazen pri osebni in strokovni rabi, zlasti v raziskovalnem delu. Agenti UI se uporabljajo kot podpora pri iskanju, združevanju in analizi informacij iz več virov. Namesto ročnega in časovno potratnega pregledovanja literature lahko ti sistemi samostojno pregledujejo dostopne podatkovne baze, presojajo ustreznost in zanesljivost virov ter iz zbranih podatkov oblikujejo strukturirane povzetke ali poročila. S tem omogočajo uporabniku, da se osredotoči na interpretacijo, kritično presojo in sprejemanje odločitev, ne pa na samo zbiranje podatkov. Takšna raba postopno vodi v oblikovanje prilagodljivih podpornih sistemov, ki se učijo uporabnikovih preferenc in raziskovalnih navad.[10]

Na področju razvoja programske opreme agenti UI presegajo vlogo orodij za dopolnjevanje ali preverjanje kode. Sposobni so analizirati obstoječe programske rešitve, prepoznati logične ali strukturne napake, izvajati teste ter predlagati ali uvesti izboljšave. Nekateri sistemi razumejo cilje projekta in

arhitekturo programske rešitve ter lahko samostojno razvijejo posamezne module ali posodobijo zastarele dele kode. Primeri, kot sta AutoGPT ali Devin, kažejo razvoj agentov, ki so sposobni večzaporednega sklepanja, ponavljajočega preverjanja rešitev in postopnega približevanja zastavljenemu cilju. Tak pristop zmanjšuje obremenitev razvijalcev pri rutinskih nalogah ter omogoča večjo osredotočenost na načrtovanje in razvoj kompleksnejših rešitev.[7,10]

Podobni učinki so opazni tudi v poslovnih procesih. Agenti UI se uporabljajo za spremljanje ključnih kazalnikov, pripravo poročil in podporo strateškemu odločanju. V podpori uporabnikom lahko samostojno obravnavajo pogoste poizvedbe in težave, v analitičnih okoljih pa sproti spremljajo trende ter prepoznavajo odstopanja ali potencialna tveganja. Njihova prednost je neprekinjeno delovanje ter sposobnost sprotnega prilagajanja novim podatkom in okoliščinam, kar prispeva k hitrejšemu odločanju in večji učinkovitosti organizacij.[11,12]

V prihodnosti se pričakuje širitev uporabe agentov UI na bolj kompleksne in medsebojno povezane naloge. Takšni sistemi bi lahko usklajevali več vidikov posameznikovega ali organizacijskega delovanja, od upravljanja urnikov in virov do podpore pri finančnem načrtovanju. Z razvojem teh tehnologij se bo meja med orodji za podporo odločanju in sistemi, ki samostojno izvajajo določene naloge, postopno zabrisala. Agenti UI bodo tako vse pogosteje delovali kot samostojni podporni sistemi, ki dopolnjujejo človeško presojo v osebnem in poklicnem okolju.[7,10,12]

Agenti UI v zdravstvu

V zdravstvenem okolju agenti UI predstavljajo podporna orodja, namenjena razbremenitvi zdravstvenih delavcev ter izboljšanju učinkovitosti in varnosti obravnave. Njihova uporaba se razteza od avtomatizacije administrativnih opravil do podpore kliničnemu odločanju, organizacije dela in raziskovalne dejavnosti.

Pomembno področje uporabe agentov UI je upravljanje zdravstvene dokumentacije. Sistemi lahko avtomatizirajo pripravo osnutkov različnih administrativnih dokumentov, kot so interni bolnišnični obrazci, postopki eNaročanja, eRecepti, odpustna pisma in povzetki obiska bolnika. Na podlagi strukturiranih in nestrukturiranih podatkov iz zdravstvene dokumentacije lahko agenti UI pripravijo predloge besedil, ki jih zdravstveni delavec pregleda in potrdi. Takšen pristop zmanjšuje časovno obremenitev, hkrati pa zmanjšuje možnost napak, ki lahko nastanejo zaradi preobremenjenosti ali nepreglednosti velike količine podatkov.[13,14]

Poleg administrativne podpore imajo agenti UI pomembno vlogo pri kliničnem odločanju. Z analizo anamneze, simptomov, laboratorijskih izvidov, slikovne diagnostike in kliničnih smernic lahko predlagajo možne diferencialne diagnoze, opozarjajo na potencialno nevarne interakcije zdravil ter priporočajo dodatne diagnostične postopke. Taki sistemi se lahko uporabljajo tudi kot orodja za zgodnje opozarjanje pri tveganih stanjih, kjer pravočasno zaznavanje odstopanj pomembno vpliva na potek zdravljenja.[13,14]

Znatna dodana vrednost agentov UI se kaže pri obdelavi in interpretaciji medicinskih podatkov. Samodejni pregled laboratorijskih izvidov omogoča hitrejše prepoznavanje patoloških odstopanj ter spremljanje sprememb vrednosti skozi čas. Podobno lahko agenti UI sodelujejo pri analizi elektrokardiogramov (EKG), računsko-tomografskih (CT), magnetno-resonančnih (MR) ali rentgenskih (RTG) slik ter pripravijo predloge izvidov ali poudarijo ključne ugotovitve, ki zdravniku pomagajo pri hitrejšem in bolj informiranem odločanju. Takšna podpora prispeva k večji natančnosti in varnosti obravnave.[14,15]

Na področju komunikacije s pacienti se agenti UI uporabljajo kot orodja za predhodno triažo in zbiranje osnovnih podatkov pred obiskom zdravstvene ustanove. Lahko strukturirano zberejo informacije o simptomih, poteku bolezni in zdravljenju ter pacientom na razumljiv način pojasnijo navodila za terapijo, spremljajo upoštevanje zdravljenja in pošiljajo opomnike za jemanje zdravil ali kontrolne preglede. S tem se razbremeni zdravstveno osebje pri rutinskih pojasnilih in osnovnih administrativnih nalogah, hkrati pa se izboljša kontinuiteta oskrbe.[14]

Agenti UI imajo pomembno vlogo tudi na organizacijski ravni zdravstvenih ustanov. Uporabljajo se za analizo zasedenosti ambulant in bolnišničnih oddelkov, optimizacijo razporejanja terminov, razvrščanje pacientov glede na triažne kriterije ter napovedovanje delovne obremenitve zdravstvenega osebja. Takšni sistemi omogočajo vodstvom boljši pregled nad delovnimi procesi in podpirajo učinkovitejše upravljanje virov.[13]

V raziskovalnem delu agenti UI omogočajo hiter pregled najnovejših znanstvenih objav, pripravo povzetkov člankov ter preverjanje skladnosti kliničnih odločitev z veljavnimi smernicami. Prav tako lahko pomagajo pri iskanju ustreznih kliničnih študij za specifične profile pacientov. Pri spremljanju

kroničnih bolnikov lahko analizirajo podatke iz nosljivih naprav, zaznajo zgodnje znake poslabšanja in pravočasno opozorijo zdravstveno osebje na potencialno tveganje.[13-15]

Pomemben vidik uporabe agentov UI v zdravstvu je tudi izboljšanje varnosti obravnave. Sistemi lahko delujejo kot dodatni nadzorni mehanizem, ki opozarja na neustrezne odmerke zdravil, manjkajoče podatke ali potencialno sporne odločitve, še preden so te dokončno potrjene. Na ta način agenti UI delujejo kot inteligenten varnostni filter, ki dopolnjuje klinično presojo in prispeva k večji kakovosti zdravstvene oskrbe.[13-15]

Primeri agentov UI v praksi

Agenti UI so danes prisotni na številnih področjih in se razlikujejo glede na svojo funkcijo, stopnjo samostojnosti ter področje uporabe. Zaradi boljše preglednosti jih je smiselno razvrstiti v več skupin, pri čemer posamezni agenti pogosto združujejo lastnosti več kategorij. V nadaljevanju so predstavljeni izbrani primeri agentov UI, razporejeni glede na njihovo osnovno vlogo in način delovanja.

Agenti za avtomatizacijo nalog

Namenjeni so izvajanju ponavljajočih se ali rutinskih opravil z minimalnim ali brez neposrednega človeškega nadzora. Med splošnimi primeri takšnih agentov sta AutoGPT[16] in BabyAGI,[17] ki omogočata večkratno zaporedno izvajanje nalog na podlagi zastavljenega cilja. V zdravstvenem okolju se agenti za avtomatizacijo uporabljajo predvsem za pripravo dokumentacije, analizo upravnih podatkov in predhodno triažo pacientov. Primer takšnega sistema je Nuance DAX Copilot[18], ki med kliničnim pregledom samodejno beleži in oblikuje zdravniške zapiske, ter Infermedica Triage AI[19], namenjen strukturiranemu zbiranju simptomov in osnovne anamneze pred obiskom zdravstvene ustanove.

Robotski in fizični agenti

Delujejo v fizičnem okolju in so sposobni zaznavanja, gibanja ter interakcije z okolico. Med splošnimi primeri takšnih sistemov sta ROSA (NASA Robot Operating System Agent)[20] in Tesla Autopilot oziroma Full Self-Driving (FSD).[21] V zdravstvu se fizični agenti uporabljajo predvsem v obliki robotskih sistemov. Da Vinci Surgical System[22] je polavtonomen kirurški sistem, namenjen minimalno invazivnim posegom, kjer robot deluje kot podpora kirurgu. Moxi[23] je bolnišnični robot, ki opravlja logistične naloge, kot so transport materiala, zdravil ali laboratorijskih vzorcev, in s tem razbremenjuje zdravstveno osebje.

Agent koder

Specializiran je za izvajanje nalog, povezanih s pisanjem, preverjanjem in testiranjem programske kode. Takšni agenti delujejo znotraj jasno opredeljenih okvirov in niso namenjeni usklajevanju drugih agentov UI. Med znanimi primeri sta OpenAI o1[24] in Devin (Cognition AI).[25] V zdravstvenem kontekstu se podobni pristopi uporabljajo pri razvoju in preverjanju kliničnih algoritmov, na primer z orodji Google MedLM Tools,[26] ki omogočajo preverjanje medicinske logike in podporo razvoju kliničnih aplikacij.

Neopredeljeni oziroma splošni agenti UI

Združujejo več funkcij, kot so pogovor, ustvarjanje besedil in osnovna analiza podatkov, brez ozke specializacije. Med splošnimi primeri so Pi.AI[27] ter komercialni pogovorni sistemi, kot so Replika, Character.AI in Samantha. V zdravstvenem okolju je bil znan primer IBM Watson for Health, ki je združeval obdelavo naravnega jezika, podporo kliničnemu odločanju in upravljanje medicinskih podatkov, a projekt je iz strokovnih in poslovnih razlogov propadel, IBM pa zdaj kot splošno platformo razvija watsonx.[28]

Infrastrukturni agenti

Ne izvajajo neposrednih nalog na vsebinski ravni, temveč zagotavljajo podporno okolje za delovanje drugih agentov. Njihova vloga vključuje upravljanje podatkovnih tokov, virov in medsebojne komunikacije med sistemi. Med splošnimi primeri so LangChain Agents[29] in LlamaIndex Agents.[30] V zdravstvu so pomembni OpenHealthAgents,[31] ki omogočajo varno interoperabilnost zdravstvenih podatkov med različnimi informacijskimi sistemi v skladu s standardom FHIR.

Eksperimentalni in raziskovalni agenti

Namenjeni so preizkušanju novih pristopov k učenju, prilagajanju in sodelovanju med agenti UI. Takšni sistemi imajo pomembno vlogo pri razvoju teoretičnih osnov in zbiranju podatkov o vedenju agentov. Med primeri sta AutoGen (Microsoft Research)[32] ter LEON oziroma Voyager (Minecraft AI Agent).[33] V zdravstvenem raziskovalnem prostoru je pomemben primer Qure.ai Radiology Agents,[34] avtonomni sistem za analizo slikovne diagnostike, vključno s CT, RTG in MR preiskavami.

Etika in varnost

Z naraščajočo stopnjo samostojnosti agentov UI se krepijo tudi etični, pravni in varnostni izzivi, ki presegajo zgolj tehnične vidike njihovega razvoja in uporabe. V zdravstvu so ti izzivi še posebej izraziti, saj lahko odločitve agentov UI neposredno vplivajo na zdravje in varnost pacientov ter na klinično odgovornost zdravstvenih delavcev.[35]

Eden ključnih etičnih izzivov je upravljanje ciljev in tveganje, da pride do neusklajenosti med nameni uporabnika in dejanskim delovanjem agenta. Agenti UI so zasnovani za optimizacijo določenih ciljev, vendar lahko brez ustreznih omejitev in kontekstualnega razumevanja ti cilji vodijo do neželenih ali celo škodljivih posledic. Tudi navidezno preprosti cilji, kot so učinkovitost, zmanjšanje stroškov ali pospešitev procesov, lahko v zdravstvenem okolju povzročijo zanemarjanje varnostnih postopkov ali kliničnih standardov, če niso ustrezno uravnoteženi z etičnimi in strokovnimi načeli.[35,36] Zato je nujno razvijati sisteme, ki vključujejo mehanizme za spoštovanje človeških vrednot, jasne omejitve delovanja ter možnost varnega in takojšnjega človeškega posega ali zaustavitve sistema.[36,37]

Pomembno področje etične razprave predstavlja tudi vprašanje odgovornosti. V primeru napake agenta UI je pogosto nejasno, ali odgovornost nosi razvijalec sistema, zdravstvena ustanova, posamezni uporabnik ali kombinacija vseh navedenih. V zdravstvenem kontekstu so takšne nejasnosti posebej problematične, saj lahko napačni predlogi ali odločitve vplivajo na klinično obravnavo pacienta. Obstoječi pravni okviri večinoma niso prilagojeni porazdeljeni odgovornosti, ki jo prinaša uporaba avtonomnih sistemov, zato se pojavlja potreba po vzpostavitvi jasnih revizijskih sledi, dokumentiranju odločitev in transparentnih mehanizmih upravljanja, ki omogočajo razumevanje, kako in zakaj je do določene odločitve prišlo.[38-40]

Preglednost delovanja agentov UI je tesno povezana z varnostjo in zaupanjem. Sistemi, ki temeljijo na kompleksnih modelih, pogosto delujejo kot t. i. črne skrinjice, kar otežuje razlago njihovih priporočil. V zdravstvu je razložljivost ključna, saj morajo zdravstveni delavci razumeti podlago predlaganih odločitev, da jih lahko ustrezno presodijo in prevzamejo odgovornost za končno klinično odločitev. Vzpostavitev razložljivih in preverljivih modelov ter jasna opredelitev vloge agenta UI kot podpornega orodja, ne pa nadomestila za klinično presojo, predstavlja temelj varne uporabe teh sistemov.[38,40]

Poleg neposrednih kliničnih tveganj imajo agenti UI tudi širše učinke na organizacijo zdravstvenega dela. Avtomatizacija administrativnih in analitičnih nalog lahko izboljša učinkovitost delovnih procesov, hkrati pa vpliva na preoblikovanje delovnih vlog in zahteve po novih znanjih zdravstvenih delavcev.[40,41] *Ključen izziv bo zagotoviti, da tehnološki napredek spremljajo ustrezni izobraževalni in organizacijski ukrepi, ki ohranjajo kakovost in varnost zdravstvene oskrbe.*[40]

Uvajanje agentov UI v zdravstvo zato zahteva celovit pristop, ki vključuje jasno regulativno okolje, standarde kakovosti, stalno spremljanje učinkov na klinično prakso ter dosledno ohranjanje človeškega nadzora. Le s takšnim pristopom je mogoče zagotoviti, da bodo agenti UI delovali kot varna in etično sprejemljiva podpora zdravstvenim delavcem, ne pa kot vir dodatnih tveganj.[35-41]

Zaključek

Agenti UI predstavljajo pomemben razvojni korak v uporabi umetne inteligence, saj omogočajo avtonomno izvajanje kompleksnih nalog v dinamičnih in podatkovno bogatih okoljih. V primerjavi s tradicionalnimi sistemi, ki se odzivajo zgolj na posamezne ukaze, agenti UI delujejo zaporedno, se prilagajajo novim informacijam in samostojno usklajujejo več nalog hkrati. V zdravstvenem okolju takšni sistemi ponujajo priložnosti za izboljšanje učinkovitosti delovnih procesov, podporo kliničnemu odločanju, boljše upravljanje podatkov ter razbremenitev zdravstvenih delavcev pri administrativnih in rutinskih opravilih.

Hkrati uporaba agentov UI v zdravstvu odpira številna etična, pravna in varnostna vprašanja, ki zahtevajo premišljen in odgovoren pristop. Ključni izzivi vključujejo zagotavljanje preglednosti delovanja, jasno opredelitev odgovornosti, ohranjanje človeškega nadzora ter zaščito pacientov pred morebitnimi tveganji, povezanimi z avtonomnim odločanjem. Brez ustreznih regulativnih okvirov, standardov kakovosti in stalnega spremljanja vplivov na klinično prakso lahko uvedba takšnih sistemov vodi do neželenih posledic.

Prihodnja vloga agentov umetne inteligence v zdravstvu bo zato odvisna od sposobnosti usklajevanja tehnološkega razvoja z etičnimi načeli, zakonodajo in strokovnimi standardi. Le ob jasno opredeljeni podporni vlogi teh sistemov, ob doslednem človeškem nadzoru in odgovorni rabi

lahko agenti UI prispevajo k varnejši, učinkovitejši in bolj trajnostni zdravstveni oskrbi.

Reference

1. Mangiò F, Pedeliento G, Wassler P, Williams N: Discursively negotiating AI: A social representation theory approach to LLM-based chatbots. *Technol Forecast Soc Change* 2025; 221: 124352. https://doi.org/10.1016/j.techfore.2025.124352
2. Matarazzo A, Torlone R: A survey on large language models with some insights on their capabilities and limitations. *arXiv* 2025. https://doi.org/10.48550/arXiv.2501.04040
3. Pasternak G, Rajagopal D, White J, Atreja D, Thomas M, Hurn-Maloney G *et al.*: Beyond reactivity: Measuring proactive problem solving in LLM agents. *arXiv* 2025. https://doi.org/10.48550/arXiv.2510.19771
4. Cheng Y, Zhang C, Zhang Z, Meng X, Hong S, Li W *et al.*: Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv* 2024. https://doi.org/10.48550/arXiv.2401.03428
5. Deng Y, Liao L, Zheng Z, Yang GH, Chua TS: Towards human-centered proactive conversational agents. *arXiv* 2024. https://doi.org/10.48550/arXiv.2404.12670
6. Zhu F, Simmons R: Bootstrapping cognitive agents with a large language model. *Proc AAAI Conf Artif Intell* 2024; 38(1): 655-663. https://doi.org/10.1609/aaai.v38i1.27822
7. Zhang R, Liu G, Liu Y, Zhao C, Wang J, Xu Y *et al.*: Toward edge general intelligence with agentic AI and agentification: Concepts, technologies, and future directions. *arXiv* 2025. https://doi.org/10.48550/arXiv.2508.18725
8. Ruan J, Chen Y, Zhang B, Xu Z, Bao T, Du G *et al.*: TPTU: Large language model-based AI agents for task planning and tool usage. *arXiv* 2023. https://doi.org/10.48550/arXiv.2308.03427
9. Bilal A, Mohsin MA, Umer M, Bangash MAK, Jamshed MA: Meta-thinking in LLMs via multi-agent reinforcement learning: A survey. *arXiv* 2025. https://doi.org/10.48550/arXiv.2504.14520
10. Jennings NR: On agent-based software engineering. *Artif Intell* 2000; 117(2): 277-296. https://doi.org/10.1016/S0004-3702(99)00107-1
11. Luo J, Zhang W, Yuan Y, Zhao Y, Yang J, Gu Y *et al.*: Large language model agent: A survey on methodology, applications and challenges. *arXiv* 2025. https://doi.org/10.48550/arXiv.2503.21460
12. Hellert T, Bertwistle D, Leemann SC, Sulc A, Venturini M: Agentic AI for multi-stage physics experiments at a large-scale user facility particle accelerator. *arXiv* 2025. https://doi.org/10.48550/arXiv.2509.17255
13. Karunanayake N: Next-generation agentic AI for transforming healthcare. *Inform Health* 2025; 2(2): 73-83. https://doi.org/10.1016/j.infoh.2025.03.001
14. Bracken A, Reilly C, Feeley A, Sheehan E, Merghani K, Feeley I: Artificial intelligence (AI) – powered documentation systems in healthcare: A systematic review. *J Med Syst* 2025; 49(1): 28. https://doi.org/10.1007/s10916-025-02157-4
15. Gabrielli D, Prenkaj B, Velardi P, Faralli S: AI on the pulse: Real-time health anomaly detection with wearable and ambient intelligence. *arXiv* 2025. https://doi.org/10.48550/arXiv.2508.03436
16. Significant Gravitas: *AutoGPT*. https://agpt.co (11. 2. 2026)
17. BabyAGI: *BabyAGI*. https://babyagi.org (11. 2. 2026)
18. Nuance Communications: *Move beyond scribes to automatically document care*. Burlington, MA 2025: Nuance Communications, Inc. https://www.nuance.com/healthcare/dragon-ai-clinical-solutions/dax-copilot/infographic/move-beyond-scribes-to-automatically-document-care.html (11. 2. 2026)
19. Infermedica: *Virtual triage – enhance symptom assessment with AI-powered virtual triage*. Denver, CO 2025: Infermedica. https://infermedica.com/solutions/triage (11. 2. 2026)
20. NASA Jet Propulsion Laboratory (JPL): *ROSA: Robot Operating System Agent – Wiki*. La Cañada Flintridge, CA 2025: NASA/JPL. https://github.com/nasa-jpl/rosa/wiki (11. 2. 2026)
21. Tesla: *Full Self-Driving (FSD) – Supervised*. Austin, TX 2025: Tesla, Inc. https://www.tesla.com/fsd (11. 2. 2026)
22. Fatima S: *Da Vinci Xi surgical robot perform kidney surgery in Dubai*. https://www.siasat.com/da-vinci-xi-surgical-robot-perform-kidney-surgery-in-dubai-2330469 (11. 2. 2026)
23. Diligent Robotics: *Moxi*. Austin, TX 2025: Dilligent Robotics. https://www.diligentrobots.com/moxi (11. 2. 2026)
24. OpenAI Group PBC: *OpenAI o1*. San Francisco, CA 2025: OpenAI Group PBC. https://openai.com/o1 (11. 2. 2026)
25. Cognition: *Devin, the AI Software Engineer*. San Francisco, CA 2025: Cognition. https://devin.ai/ (11. 2. 2026)
26. Google LLC. *MedLM models overview*. Mountain View, CA 2026: Google LLC. https://cloud.google.com/vertex-ai/generative-ai/docs/medlm/overview (11. 2. 2026)
27. Inflection AI: *Pi – your personal AI*. Palo Alto, CA 2025: Inflection AI, Inc. https://pi.ai/ (11. 2. 2026)
28. IBM: *watsonx*. Armonk, NY 2025: IBM Coorp. https://www.ibm.com/products/watsonx (11. 2. 2026)
29. LangChain: *LangChain overview*. SanFrancisco, CA 2025: LangChain. https://python.langchain.com/docs/concepts/agents/ (11. 2. 2026)
30. LlamaIndex: *Agents: LlamaIndex Python Documentation*. San Francisco, CA 2025: LlamaIndex. https://developers.llamaindex.ai/python/framework/use_cases/agents/ (11. 2. 2026)
31. Open Health Agents: *Open Health Agents*. Boston, MA in Hyderabad, Telangana 2025: Open Health Agents. https://www.openhealthagents.org/ (11. 2. 2026)

32. Microsoft: *AutoGen: A framework for building AI agents and applications*. Redmond, WA 2024: Microsoft Corp. https://microsoft.github.io/autogen/stable/ (11. 2. 2026)
33. Wang G, Xie Y, Jiang Y, Mandlekar A, Xiao C, Zhu Y, Fan L, Anandkumar A: Voyager: An open-ended embodied agent with large language models. *arXiv* 2023. https://doi.org/10.48550/arXiv.2305.16291
34. Qure.ai: *Radiology AI agents*. New York, NY 2023: Qure.ai Technologies Pvt Ltd. https://qure.ai/ (11. 2. 2026)
35. Mitchell M, Ghosh A, Luccioni AS, Pistilli G: Fully autonomous AI agents should not be developed. *arXiv* 2025. https://doi.org/10.48550/arXiv.2502.02649 (11. 2. 2026)
36. Osman N, d'Inverno M: A computational framework of human values for ethical AI. *arXiv* 2023. https://doi.org/10.48550/arXiv.2305.02748
37. Gabriel I: Artificial intelligence, values, and alignment. *Minds Mach* 2020; 30(3): 411-437. https://doi.org/10.1007/s11023-020-09539-2
38. Nguyen LH, Lins S, Renner M, Sunyaev A: *Unraveling the nuances of AI accountability: A synthesis of dimensions across disciplines*. Karlsruhe 2024: Karlsruher Institut für Technologie. https://doi.org/10.5445/IR/1000170105
39. Herbosch M: Liability for AI agents. *N C J Law Technol* 2025; 26(3): 391-458. http://dx.doi.org/10.2139/ssrn.5236649
40. Stiefenhofer P: Artificial general intelligence and the end of human employment: The need to renegotiate the social contract. *arXiv* 2025. https://doi.org/10.48550/arXiv.2502.07050
41. Shao Y, Zope H, Jiang Y, Pei J, Nguyen D, Brynjolfsson E *et al.*: Future of work with AI agents: Auditing automation and augmentation potential across the U.S. workforce. *arXiv* 2025. https://doi.org/10.48550/arXiv.2506.06576