
Pavle VALERJEV

POVIJEST I PERSPEKTIVA
RAZVOJA UMJETNE
INTELIGENCIJE U
ISTRAŽIVANJU UMA

ULOGA I PERSPEKTIVA RAZVOJA UMJETNE INTELIGENCIJE U ISTRAŽIVANJU UMA

Što je umjetna inteligencija?

AI

Ovo su dva samoglasnika koja stoje kao kratica za pojam *artificial intelligence*. Danas im se pridaje monumentalno značenje. Oni više nisu puka kratica, nego važan simbol u suvremenom društvu. Gotovo ih se može zamisliti kako stoje uklesani u mramoru kao simbol novoga doba. Takvim su ih uglavnom učinili mediji zabave popularne kulture, poput televizije i filmova, mistificirajući ta dva slova. Ali, moramo priznati, i samo područje – ljudski um – u koje zadire ova disciplina više je nego intrigantno. Naime, grubo govoreći, ona se bavi stvaranjem umjetnih umova. Ne čini li nam se da znanost opet i iznova potvrđuje svoju ulogu Prometeja, navlačeći na sebe bijes konzervativnih institucija i pojedinaca?

To je, pretpostavljam, ono što bi se u popularnom tekstu moglo ukratko reći o umjetnoj inteligenciji. No što je doista umjetna inteligencija? Sam odgovor nije jednostavan, jer definicija apstraktnih pojmova, poput onoga *inteligencija*, nikad nije bila laka. Svakom psihologu poznato je kako je u 20. stoljeću gotovo svaki istraživač inteligencije nudio svoju definiciju, a te su definicije varirale ovisno o paradigmi i rezultatima koje je istraživač dobio. Dakle, slična situacija mogla bi se očekivati i u definiranju pojma *umjetna inteligencija*. Pitamo se je li nam jednoznačna definicija doista potrebna. Konačno, sama potreba da se sve definira zapravo je arhaizam u znanosti i stoga nije ni nužna. Ipak, istraživači vole odrediti ili opisati područje kojim se bave, pa tako i istraživači umjetne inteligencije.

Rusell i Norvig (1996.) nude sljedeću definiciju: „Umjetna je inteligencija područje istraživanja i stvaranja strojeva sposobnih za takvu vrstu aktivnosti, koja bi, da su je izveli ljudi, bila proglašena inteligentnom.“ Ova je definicija prihvatljiva, no ne nudi svu šarolikost koju pojam umjetna inteligencija uključuje. Rusell i Norvig (1996.) skupili su na jednom mjestu veći broj definicija umjetne inteligencije, od kojih su mnoge slične navedenoj. No zbog

razlika u detaljima koji se tiču tipa i uzora traženoga djelovanja uočili su kako ih mogu klasificirati u četiri kategorije. Te četiri kategorije ujedno određuju i četiri glavna cilja umjetne inteligencije:

- 1) Djelovanje poput ljudi: pristup Turingova testa
- 2) Mišljenje poput ljudi: pristup kognitivnoga modeliranja
- 3) Racionalno mišljenje: pristup zakona mišljenja
- 4) Racionalno djelovanje: pristup racionalnih agenata

Djelovanje poput ljudi: pristup Turingova testa

Turingov test zove se tako po britanskom matematičaru Alanu Turingu koji je u svojem klasičnom radu (Turing, 1950.) predložio operacionalnu definiciju umjetne inteligencije. Prema Turingu, stroj je inteligentan ako osoba koja komunicira preko računalnoga terminala ne može razlučiti je li njezin sugovornik s drugoga terminala stvarna osoba ili stroj. Kako vidimo, ovdje treba razviti takav sustav koji uspješno *oponaša* ljudsku komunikaciju tako da uspješno zavara našeg ispitanika. Da bi to uspješno obavio, takav sustav mora raspolagati sposobnostima kakve su procesiranje prirodnoga jezika, reprezentacija znanja, automatsko rasuđivanje i strojno učenje. No uvijek ostaje mogućnost da se podložne reprezentacije i rasuđivanje takva sustava ne temelje na ljudskom modelu. Sam Turingov test postao je prilično popularan u računalnoj zajednici, tako da se svake godine organiziraju natjecanja s primamljivim novčanim nagradama za programere koji razvijaju programe koji će se najbolje plasirati na Turingovu testu.

Mišljenje poput ljudi: pristup kognitivnoga modeliranja

Drugi cilj umjetne inteligencije bavi se razvojem takvih sustava koji ne djeluju samo površinski – dakle na ponašajnoj razini kao kod ljudi – nego kojima i unutarnji procesi – dakle reprezentacije, algoritmi obradbe informacija – odgovaraju onima u ljudskim umovima. U ovom slučaju umjetna je inteligencija alat koji služi za modeliranje uma, tj. za testiranje teorija uma, i pri tome tijesno surađuje sa psihologijskim eksperimentiranjem, neuroznanostima, lingvistikom, filozofijom uma itd. Krovna znanost koja obuhvaća sve ove „podznanosti“ zove se *kognitivna znanost*. Kognitivna znanost je znanost o inteligentnim sustavima i njezin je temeljni zadatak postavljanje uspješne teorije uma. Dakako, te su teorije često ograničene domnom konkretnoga problema, no s druge strane dovoljno su detaljne da se mogu opisati algoritamski, a to znači i da se

moгу uključiti u računalni program. Jedna od prvih takvih teorija koja je nastala suradnjom psihologijskog eksperimentiranja i kognitivnoga modeliranja, tj. stvaranja AI sustava jest, GPS (*General Problem Solver – Opći rješavač problema*), teorija čiju su prvu verziju razvili Newell i Simon (1972.) početkom 60-ih i koja je objašnjavala strategije koje ljudi rabe u rješavanju problema. Pri tome je GPS rješavao probleme koristeći se algoritmima kojima se koriste ljudi što je rezultiralo i odgovarajućim protokolima koji su pokazali visok stupanj poklapanja s ljudskima.

Racionalno mišljenje: pristup zakona mišljenja

Sustavi koji donose zaključke racionalno se nazivaju i sustavi automatskoga rasuđivanja. Oni se temelje na nekom logičkom sustavu. Primijetite da ovdje termin „racionalno“ označuje pristup koji nije „ljudski“. Naime, ljudi nisu racionalni sustavi niti su dizajnirani da djeluju potpuno racionalno. Programi koji djeluju racionalno dizajnirani su tako da djeluju što djelotvornije, da izbjegavaju pogreške i da se koriste što štedljivijim algoritmima. Primjer idealnih racionalnih sustava jesu logički sustavi. Takvih logičkih sustava ima mnogo. Osim predikatnoga i propozicijskoga računa, tu su i modalna logika, logika s više istinosnih vrijednosti, temporalna logika, logika višega reda, teorija vjerojatnosti, teorija određenosti i meka (*fuzzy*) logika. Programi koji se temelje na racionalnim sustavima uvijek mogu, ako imaju dovoljno vremena i memorije, pronaći rješenje problema (koji se može zadati u određenom logičkom sustavu), ako to rješenje postoji. Prednost ovoga tzv. logicističkoga pristupa u umjetnoj inteligenciji nesumnjivo je u tome što se svaki logički sustav može svesti na vrlo sažet skup pravila. No glavni je nedostatak ovoga pristupa taj što često ne funkcionira u realnim situacijama, osobito kad znanje na temelju kojeg treba donijeti odluke nije potpuno ili je dvosmisleno. Nadalje, čak i samo rješenje, ako se ne može izvesti u realnom vremenu realne situacije, nego zahtijeva puno vremena, nije korisno rješenje.

Racionalno djelovanje: pristup racionalnih agenata

Prema Russellu i Norvigu (1996.), *agent* je nešto što percipira i djeluje. Dakle, ovaj pristup bavi se djelovanjem takvih agenata. Ovo djelovanje, opet, ne mora biti slično ljudskom djelovanju, ono treba biti racionalno. Racionalno djelovanje je djelovanje usmjereno prema optimalnom postizanju određenih ciljeva. Kako je u njihov opis nuž-

no uključeno djelovanje, i to po mogućnosti u stvarnim situacijama, jasno je da se ovdje inženjeri moraju nositi s problemima koje donose takve situacije. Važnost ovoga pristupa dolazi do izražaja u situacijama kad se ne može napraviti najracionalniji izbor, ali se ipak nešto mora napraviti. Na primjer, postoje situacije kada se problem mora riješiti u ograničenom vremenu. Zadatak stroja više nije da nađe najbolje rješenje, kao u prethodnom pristupu, jer za njega najčešće nema vremena, prostora, resursa, energije ili nečega poput toga. To su realne situacije. Od stroja se, dakle, traži da nađe dovoljno dobro rješenje ili čak i *bilo kakvo rješenje* (koje je bolje od nikakva rješenja) te da *djeluje*. Ova vrsta racionalnosti zove se *ograničena* racionalnost. Dakle, kako ovaj pristup često zahtijeva neku vrstu djelovanja u fizičkoj okolini, jasno je da je on vrlo važan za područje robotike.

Postojanje ovih četiriju ciljeva umjetne inteligencije govori o tome kako svi istraživači ne doživljavaju podjednako ulogu i ciljeve ove discipline. Svakako, za psihologe i kognitivne znanstvenike važnija su prva dva pristupa, iako ne isključivo. Jer, najopćenitiji cilj kognitivne znanosti nije samo objašnjenje ljudske inteligencije nego i postavljanje dovoljno općenite teorije koja može opisati svaku pojavu inteligentnoga ponašanja.

Velika raznovrsnost problema koji se javljaju pri bavljenju umjetnom inteligencijom dovela je do podjele discipline ne samo prema ciljevima nego i problemima s kojima se suočavaju istraživači, koji onda određuju posebne grane unutar discipline.

Glavne grane umjetne inteligencije

Klasifikacija problema kojima se bave istraživači umjetne inteligencije generirala je slijedeće grane te discipline:

1) *Rješavanje problema* je grana koja uključuje razvoj algoritama, tj. strategija pretraživanja problemskoga prostora, što se nerijetko naziva i heuristikom, reprezentacijom problemskoga prostora, konstruiranjem inteligentnih programa za igranje igara s diskretnim stanjima, poput šaha, a u novije vrijeme i raznih strateških igara u realnom vremenu.

2) *Reprezentacija znanja* je grana koja se bavi problemima predstavljanja i pohranjivanja korisnih informacija, izgradnje baze znanja i povezivanjem tog znanja s drugim znanjem na iskoristiv način. Ovi problemi važni su u razvoju, recimo, ekspertnih sustava.

3) *Automatsko rasuđivanje* je grana koja uključuje razvoj programa što se temelje na logičkim zakonitostima, kao što su npr. dokazivatelji teorema.

4) *Planiranje i djelovanje* je grana koja je važna u pristupu racionalnih agenata. Dobar agent može izračunati optimalan tijek prikladnih ciljeva i akcija te pravilno djelovati u skladu s planom. Ovakav agent ima sličnosti, ali i posebnosti, u odnosu na rješavače problema.

5) *Rasuđivanje u neodređenim uvjetima* je posebna kategorija problema kod koje se javljaju posebne poteškoće i kod reprezentacije neodređenoga znanja i kod odabira ispravnih algoritama za izvođenje optimalnih odluka. Temelji takvih sustava jesu teorija vjerojatnosti i teorija korisnosti, koje zajedno čine teoriju odlučivanja na kojoj se temelji izgradnja racionalnih agenata za neodređene svjetove.

6) *Učenje* – strojno je učenje nužna sposobnost bilo kojega umjetno-inteligentnog sustava. Ovisno o sustavu, razlikuju se i algoritmi koji omogućuju određen tip učenja. Tako se kod agenata može primjenjivati oblik induktivnoga učenja koji se zove stablo odluka. Nadalje, kod arhitektura koje se služe neuralnim mrežama postoje brojni algoritmi učenja, od kojih je jedan od najpoznatijih tzv. širenje unatrag (*back-propagation*). Kod tzv. *mreža vjerovanja* rabe se algoritmi bayesovskog učenja.

7) *Procesiranje prirodnoga jezika, razumijevanje govora, automatski prevoditelji* važne su grane koje bi trebale omogućiti laku komunikaciju sa strojevima. Usprkos naporu na ovim područjima, nastaju mnogi problemi povezani s problemima reprezentacije znanja.

8) *Kompjutorski vid* je grana koja se bavi problemima vidne percepcije. Stroj bi trebao moći reprezentirati i prepoznati scenu koju „vidi“ kamerom, tako da razlikuje objekte od pozadine i jedne od drugih, da može odrediti njihovu lokaciju u prostoru te da objekt može prepoznati i povezati s reprezentacijom odgovarajućeg objekta iz vlastite memorije.

9) *Robotika* je grana koja može uključivati, ovisno o predviđenim zadacima robota, sve prethodno navedene grane. U najmanju ruku mora uključivati kompjutorski vid te planiranje i djelovanje, jer robot je stroj koji se kreće i djeluje u realnoj fizičkoj okolini.

KRATKA POVIJEST UMJETNE INTELIGENCIJE

Da bi se lakše stekla slika o tome što je umjetna inteligencija danas, odnosno na kojem se stupnju razvoja nalazi i uolikoj mjeri može simulirati inteligentno ponašanje, u nastavku će biti vrlo sažeto iznesena povijest umjetne inteligencije. Dodatni razlog za iznošenje povijesti umjetne inteligencije jest taj što se to područje nije razvijalo nezavisno, već u skladu s razvojem spoznaja iz kognitivne psihologije, lingvistike i neuroznanosti.

Umjetna inteligencija ima jako dugu pretpovijest, koja se tiče znanstvenih disciplina što su svojim spoznajama utjecale na razvoj umjetne inteligencije. Ta pretpovijest seže u doba antike te razvoja filozofije i logike. Mnogi od problema teorije spoznaje koji su se javili u dugoj povijesti filozofije aktualizirani su pojavom umjetne inteligencije. Nadalje, osim filozofije, važan je i doprinos matematike, pogotovo grana koje se bave izračunljivošću, kompjutacijom, kompleksnošću i vjerojatnošću. Dakako, ne smijemo zaboraviti ni pojavu eksperimentalne, posebice kognitivne, psihologije. Svakako treba spomenuti i utjecaj računalnog inženjerstva te suvremene lingvistike u 20. stoljeću.

Radovi koji su prepoznati kao umjetna inteligencija javljaju se još 40-ih godina 20. stoljeća. Oni su se tada temeljili na tri izvora: znanju o živčanom sustavu, formalnoj analizi logike sudova Russela i Whiteheada, te Turingovoj teoriji kompjutacije. Prvi takav rad jest onaj iz 1943., u kojem su McCulloch i Pitts predložili model umjetnih „on-off“ neurona koji su povezani u mrežu i funkcioniraju po načelu „sve ili ništa“. U ranim 50-im godinama Shannon i Turing, nezavisno, pišu prve programe za šah. Minsky i Edmonds 1951. godine grade prvo računalo – neuralnu mrežu SNARC koji se sastoji od 3000 vakuumskih cijevi, a simulirala je neuralnu mrežu od 40 neurona. Zanimljivo je kako se Minskyjeva doktorska komisija dvoumila može li se taj rad smatrati matematičkim, no von Neumann koji je sjedio u komisiji proročanski je rekao: „Ako nije sada, hoće jednog dana.“ Sustav je služio kao automatski pilot za bombarder B-24. Samuel 1952. godine piše prve programe za igru dame, a Devol 1954. patentira prvu robotsku ruku.

Godina 1956. važna je za ovo područje. Na Dartmouth Collegeu, na dvomjesečnoj radionici za 10 polaznika, McCarthy je skovao termin *artificial intelligence*. Taj se događaj smatra rođenjem ovoga područja. Newell i Simon, dvojica zaslužnih pionira koji nisu bili na tom seminaru, za to su vrijeme razvijali *Logic Theorist* (LT).

To je program za automatsko rasuđivanje koji sam može izvoditi logičke teoreme. McCarthy 1958. godine razvija LISP (*List Processing*) – dominantan programski jezik za umjetnu inteligenciju. Opisuje i *Advice Taker* – hipotetski program koji je prvi cjelovit sustav umjetne inteligencije i ujedno cjelovita kognitivna teorija uma jer opisuje sposobnost reprezentacije i stjecanja novoga znanja bez reprogramiranja. Iste godine nastaju i prvi eksperimenti s genetskim algoritmima (strojnom evolucijom). Rochester 1959. godine razvija još jedan program za automatsko rasuđivanje - *Geometry Theorem Prover*.

Newell i Simon nastavljaju svoj rad i 1961. godine razvijaju važan program - *General Problem Solver* (GPS), program koji rješava probleme tako da simulira ljudske protokole koristeći se strategijom analize sredstva i cilja. To je vjerojatno prvi uspješni model ljudskoga mišljenja. Rosenblatt 1962. godine razvija PERCEPTRON – formu neuralne mreže koja se i danas upotrebljava. Na tim mrežama dokazano je kako algoritam za učenje može prilagoditi snagu veza perceptrona da se usklade s bilo kojim danim ulazom. Od 1963. do 1968. Minskyjevi studenti razvijaju niz sustava koji rješavaju probleme u ograničenim domenama – tzv. mikrosvijetovima (integrali, geometrijske analogije, algebarski problemi itd.). Weizenbaum 1965. godine razvija poznati, no ne previše uspješan, program ELIZA koji simulira terapeuta u konverzaciji s klijentom. Program se temeljio na jednostavnoj sintaktičkoj manipulaciji i imao je minimalnu mogućnost baratanja znanjem.

Razdoblje do sredine 60-ih godina bilo je vrlo optimistično i obilježeno velikim očekivanjima zbog ranih uspjeha. Stoga ga suvremeni istraživači umjetne inteligencije u šali nazivaju erom „Vidi, mama, bez ruku!“ Postoji i anegdota da je Minsky u zanosu uspjeha dao nekom svom studentu da u nekoliko mjeseci napravi sustav koji će simulirati ljudsku vidnu percepciju. No pokazalo se da je to tako složen problem da ni do danas nije riješen. Nakon toga uslijedilo je razdoblje u kojemu se uvidjelo kako su problemi stvaranja cjelovitijih inteligentnih sustava mnogo složeniji nego što se mislilo. Tada nastaju i prvi ekspertni sustavi. Tako 1969. godine Feigenbaum razvija DENDRAL – prvi sustav utemeljen na znanju (tzv. ekspertni sustav) koji daje kemijske formule spojeva na temelju rezultata spektrometrije, a 1972. godine razvija MYCIN, ekspertni sustav za medicinsku dijagnostiku. Nastaju i prvi programi za razumijevanje prirodnoga jezika. Winograd 1972. godine razvija program SHRDLU za razumijevanje prirodnoga jezika u mikrosvijetu od geometrijskih bloko-

va. Iste godine Minsky uvodi ideju okvira kao reprezentacijske sheme znanja, a Schank 1977. uvodi ideju skripta kao reprezentacijske sheme znanja za stereotipne akcije. Te su ideje preuzete iz psihologije, gdje se odgovarajuće reprezentacije nazivaju kognitivnim shemama.

U 80-im godinama 20. stoljeća umjetna inteligencija postaje industrija. Proizvode se brojni ekspertni sustavi za specifične potrebe. Na scenu se vraćaju i neuralne mreže. Taj se pristup u umjetnoj inteligenciji naziva konekcionizam, dok se onaj klasični, utemeljen na znanju, naziva simbolizam. Godine 1985. više od 180 000 robota radi na proizvodnim trakama širom svijeta i više od 100 kompanija ugrađuje sustave s robotskim viđenjem. Godine 1988. naglašava se uloga vjerojatnosti i teorije odlučivanja u umjetnoj inteligenciji. Razvijen je formalizam mreža vjerovanja za rasuđivanje pod neodređenim uvjetima. Nadalje, dolazi i do napretka u prepoznavanju govora i na tom području dominiraju skriveni Markovljevi modeli.

Od 90-ih pa do danas osjeća se sve brži napredak u robotici (autonomni roboti), kompjutorskom vidu, strojnom učenju i reprezentaciji znanja. Razvijaju se i cjelovite kognitivne arhitekture započete u 80-ima, koje pretendiraju na to da budu cjelovite teorije kognicije. Prvu takvu, ACT (*Adaptive Control of Thought*), razvio je Anderson (1983.). Poznata je takva arhitektura i SOAR, koju su razvili Newell, Laird i Rosenbloom potkraj 80-ih. Te arhitekture doživljavaju brojne revizije. Uvedena su natjecanja za programe koji prolaze Turingov test. Stvoreni su napredni šahovski programi (poput Deep Thought koji pobjeđuje svjetskoga prvaka). Ubrzan je razvoj robotike, koja se sve više komercijalizira i u industriji i u običnoj potrošnji i zabavi, no postaje nezamjenjiva u opasnim i teško dostupnim okolišima, kao što su npr. svemirska istraživanja.

Napokon, sve je više i malih aplikacija umjetne inteligencije za specifične zadatke, poput procesiranja jezika, umjetnih tutora i slično, koje se gotovo i ne primjećuju i koje sve češću primjenu imaju u svakodnevnim korisničkim programima, igrama, uređajima, osobnim vozilima i drugim mjestima. Umjetna inteligencija primjenjiva je i u uslužnim djelatnostima, npr. kao službenik za prodaju karata ili pak u informiranju o stanju u prometu.

UMJETNA INTELIGENCIJA U KOGNITIVNOJ ZNANOSTI – KOGNITIVNO MODELIRANJE

Umjetna inteligencija bavi se programiranjem računala da se ponašaju na onaj način koji bi, kad bi se promatrao kod ljudskih bića, bio proglašen inteligentnim. Ako se tim programima doda i kvaliteta da daju rješenja onako kako to rade i ljudi te da griješe na istim mjestima kao i ljudi, tada je riječ o simulaciji ljudskoga ponašanja (Simon i Kaplan, 1993.). Ako simulacija daje rezultate koji su usporedivi s rezultatima dobivenim od ljudi, tada se može zaključiti da procesi koji iskorištavaju AI program daju potencijalno valjan model tog aspekta ljudske kognicije.

Možda bi se mogao steći dojam da je računalna simulacija ljudskoga ponašanja kao metoda u kognitivnoj psihologiji poprilično različita od uobičajene metode – eksperimenta, poglavito ako kažemo: psiholozi žele proučavati ljude kao sustave obradbe informacija da objasne razne aspekte ljudske kognicije, a AI istraživači žele dizajnirati odgovarajuće algoritme za razne aspekte ljudske kognicije (Lenhart, 1984.). Ali, metodologija se može razmotriti i na slijedeći način:

Tablica 1
Usporedba metodologije
eksperimentalnoga
psihologa i AI istraživača

Koraci	Metodologija eksperimentalnog psihologa	Metodologija AI istraživača
1	Prijedlog teorije koja objašnjava fenomen	Prijedlog teorije koja objašnjava fenomen
2	Nacrt eksperimenta koji testira teoriju	Primjena teorije u računalnom programu dizajniranom da simulira fenomen
3	Obavljanje eksperimenta	Pokretanje programa
4	Analiza eksperimentalnih podataka	Analliza programskog izlaza

Ako eksperiment ne potvrđuje teoriju (tj. nije „uspješan“), tada se treba vratiti ili na korak 1, da se revidira ili preformulira teorija, ili pak na korak 2, da se revidiraju nedostaci u eksperimentalnom dizajnu. U slučaju „neuspješnog“ AI programa treba se vratiti na korak 1 radi revizije ili reformulacije teorije; ili na korak 2 da se poprave bugovi, odnosno programske pogreške.

Kada se metodologija sagleda na ovakav način, očito je da postoji paralela. Obje metodologije prihvaćaju ili odbacuju hipotezu. Tehnički problemi AI programa odgovaraju problemima pri dizajniranju eksperimenta. Zanimljiva je i sličnost u proceduralnoj petlji. Programi koji su neuspješni na zanimljiv način analogni su eksperimentima koji su neuspješni na zanimljiv način. U oba slučaja istraživač će naučiti nešto o svojoj hipotezi što ne bi uočio bez obavljanja istraživanja.

Lenhart (1984) zaključuje da, kad ovako promatramo, glavnu razliku između psihologije i AI čine oruđa koja se koriste. Ipak, postoje i druge razlike. Računalni programi mogu se kao oruđe rabiti i za opisivanje teorija stohastičke prirode, koje se ne bi mogle ubrojiti u umjetnu inteligenciju.

Relativno je lako napraviti simulaciju jednog procesa. Ali snaga modela ne počiva samo na poklapanju s empirijskim podacima određenog eksperimenta. Što je sa sličnim zadacima? Jak je model integriraniji i sadrži u sebi dovoljno široku teoriju čija računalna primjena može rješavati primjere koji nisu bili zadani u originalnom eksperimentu. Ako se dobrom modelu zadaju novi zadaci, on može dati nove odgovore, a oni se onda mogu provjeriti na ljudskim ispitanicima. Na taj način računalo nije samo alat koji pomaže da se opiše teorija nego i izvor predviđanja za buduća istraživanja. Potpuno iskorištavanje računalne simulacije zahtijeva gledanje na model na samo kao na produkt već prije kao na razvijajući entitet čije varijacije mogu sugerirati nova istraživanja. Po tome, konačno stanje u simulaciji ujedno je i početno. Jednom kad se model konstruira i testira, vrijeme je da se konstruira revizija toga modela ili različit model, i to ili sa svrhom generalizacije modela i arhitekture, primjenjujući njega i njegove varijante, ili da se testiraju nova predviđanja. Teško je otprve dobiti simulaciju koja točno predviđa podatke.

Tipovi kognitivnih arhitektura

Kod umjetne inteligencije za modeliranje kognicije dva su temeljna pristupa. To su simbolizam – ili pristup utemeljen na znanju – te konekcionizam – ili pristup neuralnih mreža.

Simbolizam se temelji na simbolima ili eksplicitnom znanju, tj. da postoji baza eksplicitnoga znanja uređenog na odgovarajući način. Sustav mora uključivati i pravila koja omogućuju pretraživanje baze znanja, izvođenje zaključaka, pohranjivanje i dozivanje znanja te uvrštavanje

novih pojmova u bazu. Ovakve arhitekture osobito su prikladne za sustave rješavanja problema kao što je GPS Newella i Simona (1972.), zatim za ekspertne sustave koji oponašaju rad stručnjaka, te za neke od cjelovitih arhitektura kognicije, kakva je ACT* što ju je razvio Anderson (1983.). Ova arhitektura predstavlja cjelovit model, tj. teoriju ljudske kognicije. Ona uključuje dugoročna memorijska spremišta za deklarativno i proceduralno znanje, koja su u međusobnoj vezi i u vezi s vanjskim svijetom preko ograničene radne memorije. Deklarativno znanje organizirano je u obliku semantičke mreže, a proceduralno u obliku produkcijskih pravila. Radno pamćenje nije samo po sebi fizički odvojena struktura, nego predstavlja onaj dio dugoročnih pamćenja koji je trenutačno aktivan. Postoje algoritmi za pretraživanje, pohranjivanje i dozivanje znanja, i oni za stvaranje novoga znanja – učenje.

Konekcionizam je pristup koji se temelji na paralelnoj i distribuiranoj kompjutaciji malih procesirajućih jedinica bliskih apstraktnom neuronu povezanih u mrežu. Temeljna strategija konekcionista pristupa jest da se uzme takva temeljna procesirajuća jedinica koja je bliska apstraktnom neuronu (Rumelhart, 1996.). Kompjutacija se tada izvodi paralelno, jednostavnim interakcijama među takvim neuronima. Ideja je da takvi procesirajući elementi komuniciraju slanjem brojevanih vrijednosti duž linija koje povezuju te elemente. Jedna od osnovnih razlika konekcionista modela od onih simboličkih jest u zahtjevu da se cjelokupno znanje nalazi u vezama među jedinicama, tako da se vrste veza razlikuju od modela do modela. Dakle, znanje je implicitno u strukturi modela koji se njime koristi, za razliku od eksplicitnoga znanja u simboličkim arhitekturama. Informacija nije određena simbolima već obrascem aktivnosti procesirajućih jedinica. Upotreba ovakvih kompjutacijskih sustava u sustavima nalik onima mozgovnima, ne nudi samo nadu da se opišu plauzibilniji načini na koje mozak izvodi zadatke obradbe informacija. Ovakav tip arhitektura nudi i jednostavnija rješenja za određene klase problema koje se vrlo teško rješavaju na klasičan, serijski način. Takvi su, na primjer, problemi najboljega slaganja ili optimizacije. Ove arhitekture najčešće se rabe za modeliranje akcije, percepcije, navigacije, motorne kontrole, pokreta tijela i drugih oblika vezivanja sa svijetom, dok su klasične arhitekture povezane s odvojenim, često deduktivnim zadacima koji se tiču raznih oblika rješavanja problema i baratanja znanjem. Nadalje, konekcionista arhitekture nastoje biti “u realnom vremenu” u smislu zahtjeva za bliskim povezivanjem trajanja kompjutacijskih procesa i trajanja subjektivne domene.

Primjer konekcionističke arhitekture jest ona koju su razvili Hummel i Biederman (1992.). To je mreža koju čini sedam slojeva procesirajućih jedinica i koja omogućuje percepciju i prepoznavanje oblika. Proces počinje ulaznim dvodimenzionalnim slojem procesirajućih jedinica, na kojemu rubovi slike aktiviraju odgovarajuće jedinice. Kroz slojeve procesirajućih jedinica obavlja se analiza osi, vertikala, volumena, određivanja geona (primitivnih geometrijskih volumena), svojstava geona, kao i relacija u kojima se ta svojstva nalaze. Konačno, gomilanje karakterističnih svojstava geona rezultira aktivacijom takvih veza koje dovode do prepoznavanja i imenovanja zadanog oblika.

Za detaljniji uvid u simbolističke arhitekture pogledati Anderson (1983.) ili Newell i sur. (1996.), a za detaljniji uvid u konekcionističke arhitekture pogledati Rumelhart (1996.).

PROBLEMI I KRITIKA AI

Usprkos brojnim uspjesima koje je postigla umjetna inteligencija, mora se prikazati i njezina tamna strana. Naime, u filozofskim krugovima zadnjih se desetljeća pokreću polemike o tome može li se ono što je do danas izvedeno smatrati inteligencijom i je li „prava“ umjetna inteligencija uopće moguća. Konačno, može li se uopće ljudski um svesti na stroj koji se može ne samo oponašati nego i stvoriti umjetnim putem? Sažeto prikazani skupovi filozofskih problema vezanih uz umjetnu inteligenciju bili bi sljedeći:

1) Problemi *kvalie* (kvalitete doživljaja) tiču se sljedećeg pitanja: što doista određuje kvalitetu doživljaja u smislu crvenila ružinih latica i plavetnila neba. *Kvalia* je toliko subjektivna stvar da filozofi dokazuju kako nikad ne bi mogli ustanoviti podudaraju li se ili razlikuju *kvalie* među različitim pojedincima. Stoga se mogu postaviti i pitanja je li moguće uspoređivanje među *kvaliama* raznih kognitivnih sustava i može li umjetna inteligencija imati kvalitetu subjektivnoga doživljaja u smislu doživljavanja crvenog, hladnog ili slatkog.

2) Drugi skup problema tiče se svijesti. Što je točno svijest i mogu li se procesi svijesti opisati kompjutacijski? Ako da, znači li to da je moguća i svjesna umjetna inteligencija? S druge strane, kognicija bez svijesti moguća je u današnjoj umjetnoj inteligenciji, no hoće li umjetna inteligencija zauvijek ostati u okviru filozofskih zombija kojima je svijest nepoznanica? Konačno, ako je kognicija

bez svijesti moguća, kakav je doprinos svijesti kogniciji? Je li ona nužno unapređenje kognicije ili je njezin puki nusprodukt i/ili epifenomen?

3) Na postavljena pitanja vežu se i problemi odnosa tijela i uma. Ti se problemi tiču načina povezivanja različitih razina organizacije kognitivnih sustava – od bioloških procesa u mreži neurona do manipulacije znanjem i mišljenja – i svode se na osnovno pitanje: *kako tijelo proizvodi um?* Nadalje, uz to se veže i pitanje zamjenjivosti wetvera (biološkoga hardvera) s neživim hardverom. Naime, ako se um može opisati algoritmima, znači li to da ga može izvoditi bilo kakav stroj, bio on mozak, silicijski stroj ili pak nešto treće?

Konačno, ostaje i pitanje dokle se najdalje može doći u razvoju umjetne inteligencije. Postoje li ograničenja, takva da se inteligencija može samo simulirati (pristup „slabe“ umjetne inteligencije), ili se mogu stvoriti cjelovita inteligentna i svjesna bića („jaka“ umjetna inteligencija)?

BUDUĆNOST UMJETNE INTELIGENCIJE

Na temelju današnjih trendova u znanosti, potreba društva te načina funkcioniranja tržišta možemo iznijeti neke pretpostavke o tome kako bi mogao izgledati razvoj umjetne inteligencije u bliskoj budućnosti. Dakle, pretpostavljamo četiri glavna smjera razvoja u idućoj dekadi.

1) Umjetna inteligencija kao opći svakodnevni proizvod: brojni programi i fizički sustavi koji su sposobni izvoditi razne manje *inteligentne* zadatke (igre, roboti za industriju, domaćinstvo i zabavu, uslužne djelatnosti, telekomunikaciju), a koje prosječni potrošači često i ne prepoznaju kao inteligentne.

2) Umjetna inteligencija u obrazovanju kao inteligentni tutori koji su sposobni optimalno, adaptivno i individualno voditi učenika kroz proces stjecanja znanja na svim područjima.

3) Umjetna inteligencija kao filter informacija. Zahvaljujući Internetu, imamo eksploziju informacija u kojima se javlja sve više neupotrebljivih informacija, odnosno šum. Dakle, suočavamo se s paradoksom po kojemu je informacija sve više, no do njih je i sve teže doći, naravno kad govorimo o korisnim informacijama. Intelligentni filteri i tražilice trebali bi pronaći informaciju koja nam doista treba te dati smislen pregled, sažetak velike količine pronađenih informacija u kratkom i funkcionalnom izvješću.

4) U znanosti i tehnologiji očekuje se sve šira upotreba ekspertnih sustava, pa čak i umjetnih znanstvenika koji su sposobni sami generirati i testirati znanstvene hipoteze na temelju dobivenih podataka. Osim toga, što se tiče samih istraživanja, u okviru umjetne inteligencije očekuje se daljnji razvitak u modeliranju ljudske kognicije i alternativnih kognicija te približavanje općoj teoriji kognicije.

5) Jak razvoj robotike za vojne potrebe te potrebe istraživanja i iskorištavanja opasnih okolina (npr. svemira). Očekuje se daljnji i velik razvoj inteligentnih letjelica, vozila, oružja koja mogu obavljati opasne i okolinski zahtjevne zadatke bez neposredne prisutnosti ljudi za koje rade.

ETIČKI I PRAVNI PROBLEMI: ŠTO AKO USPIJE?

Pod uspjehom se razumijeva potpun uspjeh: strojevi visokih razina inteligentnoga ponašanja koji mogu ravnopravno komunicirati s ljudima. Što će se onda dogoditi u društvima? Koje će biti implikacije ovakva razvoja na socijalne znanosti, ekonomiju, pravo, etiku? Premda zasad postoje male naznake u tome smjeru, javne se polemike ne bave toliko mogućnošću umjetnih inteligentnih bića. Polemike i moralna panika još su uvijek usmjerene prema genetičkim istraživanjima, pogotovo otkad je humano kloniranje postalo realna mogućnost.

Ipak, moramo se zapitati što bi se moglo dogoditi ako na našem planetu zažive visokointeligentna bića koja smo sami stvorili? Kakav će biti etički i pravni odnos prema njima? Hoće li se smatrati građanima i kojega reda? Kako će se prema tome odrediti kazneni zakon, i to za kaznena djela u kojima su umjetna bića počinitelji i za djela u kojima su žrtve? Hoće li društvena svakodnevnica biti obilježena oblikom *ugljičnoga šovinizma*, po kojemu bi se status punopravnoga bića mogao priznati samo bićima od ugljika, ali ne i umjetnim bićima? Dio filozofskih problema proizlazi iz teleološkog određenja tih bića. Postoje li ona jer su stvorena sa svrhom kojoj moraju služiti ili se mogu u jednom dijelu vlastita postojanja odrediti kao slobodna bića? Konačno, teologija i religijske institucije neće biti pošteđene vlastitih polemika: Kako riješiti probleme tvorca, duše i spasitelja umjetnih bića? Zasad rješenja takvih problema možemo pronaći uglavnom u znanstveno-fikcijskim djelima.

Umjesto zaključka može se istaknuti kako je usprkos uspjesima područje umjetne inteligencije još uvijek vrlo daleko od cjelovitih kognicija, no premda je cijela priča dobila zavidnu teorijsku i filozofijsku pozadinu, čini se da je sam pristup istraživača umjetne inteligencije najbliži onom inženjerskom: rad na rješavanju konkretnih problema i na neprekidnom malom usavršavanju postojećih sustava, bez opterećivanja snovima o stvaranju umjetnih mislećih bića. Zasad se može pretpostaviti kako će se kretati razvoj AI u bliskoj budućnosti, međutim nitko još ne može pouzdano reći gdje su krajnji dosezi ove tehnologije i kakve će sve implikacije ona imati na shvaćanje čovjeka i na funkcioniranje društva.

LITERATURA

- Anderson, J. R. (1983.), *The Architecture of Cognition*, Cambridge, MA, Harvard University Press.
- Hummel J. E. and Biederman, I. (2002.), Dynamic binding in a neural network for shape recognition. In: T. A. Polk and C. M. Seifert (Eds.), *Cognitive Modeling*, (str. 337-374), Cambridge, MA, A Bradford Book.
- Lenhart, W. G. (1984.), Paradigmatic issues in cognitive science. In: J. R. Miller, P. G. Polson, and W. Kintsch (Eds.), *Method and Tactics in Cognitive Science*, (str. 21-50), Hillsdale, NJ, Lawrence Erlbaum Associates.
- Luger, G. F. (1994.), *Cognitive Science: The Science of Intelligent Systems*. San Diego, CA, Academic Press.
- Luger, G. F. and Stubblefield, W. A. (1998.), *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Harlow, UK, Addison Wesley Longman, Inc.
- Newell, A., Rosenbloom, P. S. and Laird, J. E. (1996.), Symbolic architectures for cognition. In: M. I. Posner (Ed.), *Foundations of Cognitive Science*, (str. 93-132), Cambridge, MA, MIT Press.
- Newell, A. and Simon, H. A. (1972.), *Human Problem Solving*. Englewood Cliffs, NJ, Prentice-Hall.
- Rumelhart, D. E. (1996.), The architecture of mind: a connectionist approach. In: M. I. Posner (Ed.), *Foundations of Cognitive Science*, (str. 133-160), Cambridge, MA, MIT Press.
- Russell, S. and Norvig, P. (1996.), *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ, Prentice-Hall.
- Simon, H. A. and Kaplan, C. A. (1996.), Foundations of cognitive science. In: M. I. Posner (Ed.), *Foundations of Cognitive Science*, (str. 1-48) Cambridge, MA, A Bradford Book, MIT Press.
- Turing, A. (1950.), Computing machinery and intelligence, *Mind*, 59: 433-460.