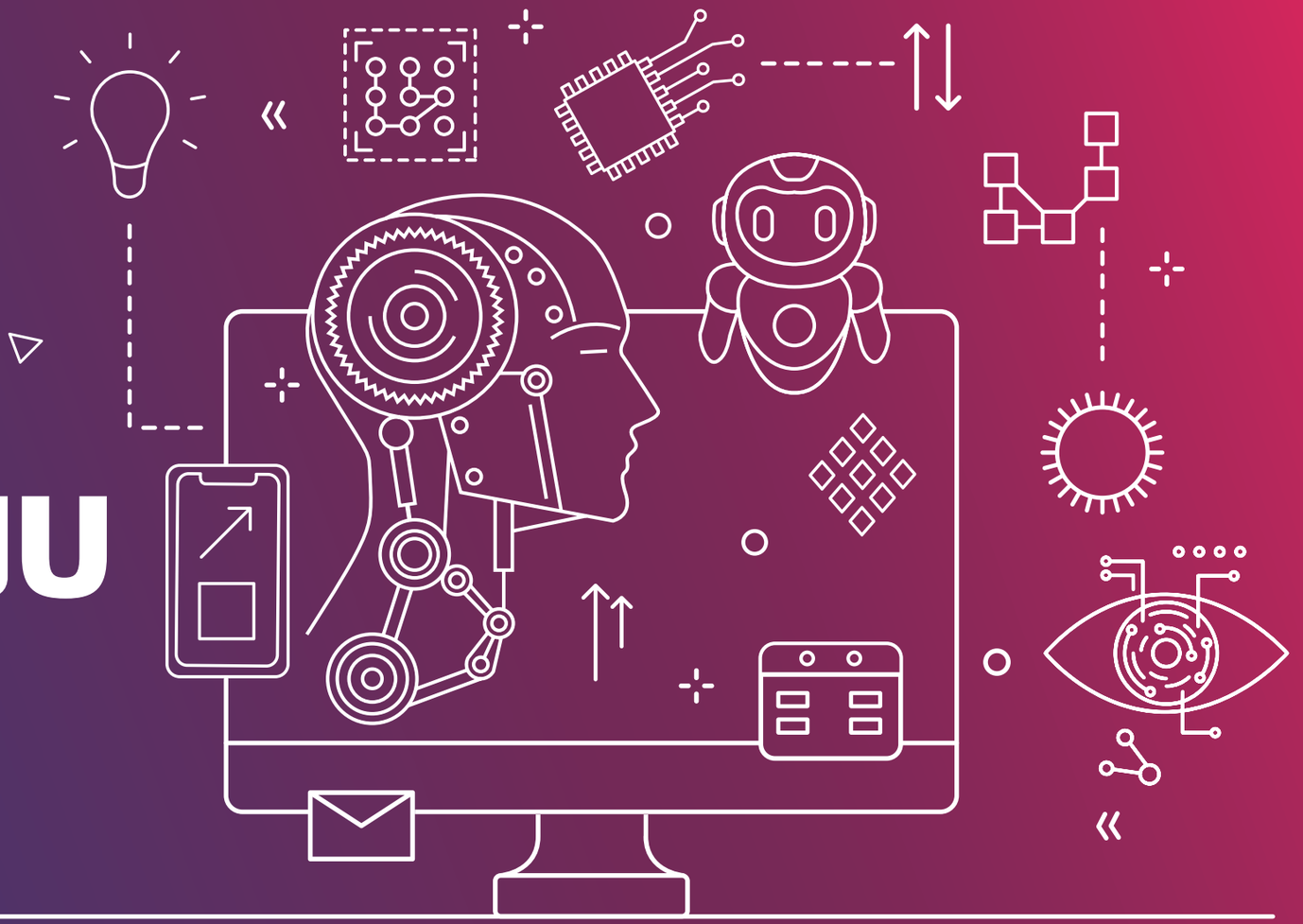
USAID
OD AMERIČKOG NARODA

Inicijativa
Digitalna Srbija

SERBIAN
AI SOCIETY

UVOD U **VEŠTAČKU INTELIGENCIJU**



Autor: **Serbian AI Society**

SADRŽAJ

Šta predstavlja AI? 3

Važnost podataka 9

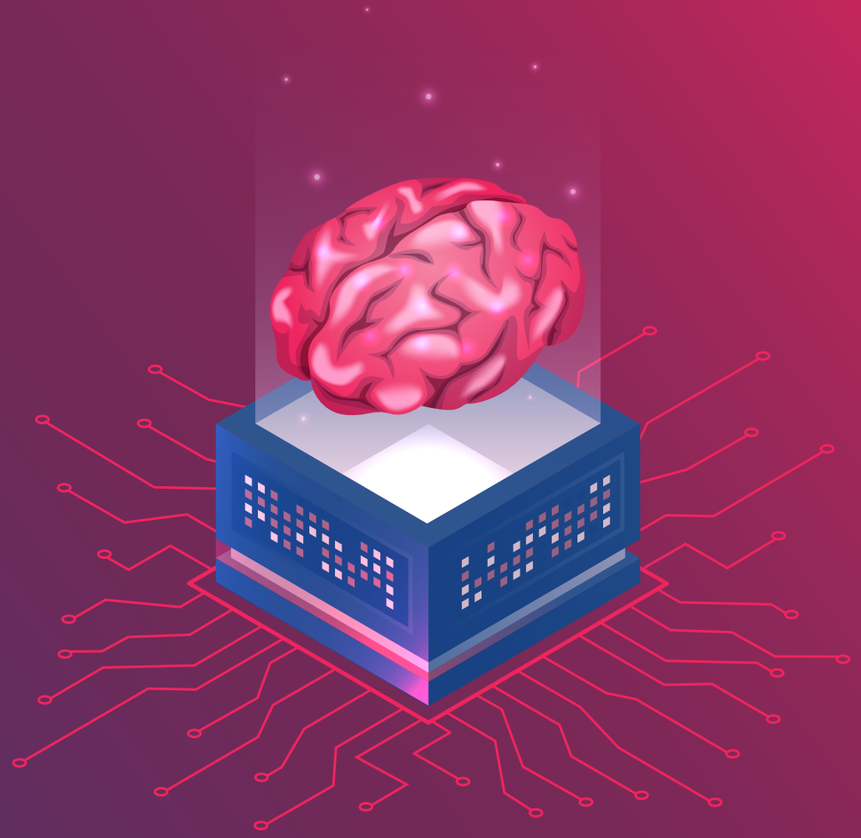
Praktični AI projekti 20

Uticaj AI na društvo 26

Dodatni resursi 30

Izrada ovog vodiča omogućena je uz podršku američkog naroda putem Američke agencije za međunarodni razvoj (USAID).
Sadržaj ovog vodiča je isključivo odgovornost Inicijative „Digitalna Srbija“ i ne predstavlja nužno stavove USAID-a ili Vlade SAD.

ŠTA PREDSTAVLJA AI?



Veštačka inteligencija (engl. *Artificial intelligence*, skraćeno AI) je sveprisutan pojam: od vesti o robotima koji igraju fudbal i pomažu u rehabilitaciji pacijenata, preko mašina poput Siri ili Alekse koje komunciraju sa nama, do filmova i knjiga naučne fantastike. Pod veštačkom inteligencijom se podrazumeva bilo koji računarski sistem koji oponaša prirodnu inteligenciju kako bi se izvršio određen zadatak.

Veštačka inteligencija se može ugrubo podeliti na dva tipa: sisteme zasnovane na pravilima i sisteme koji uče (u ovu grupu se najčešće ubraja oblast mašinskog učenja).

U automatizaciji industrijskih procesa je često dovoljno imati sistem sa unapred zadatim skupom činjenica i pravila: robotska ruka u fabrici automobila ima repetitivan zadatak koji je isti svakog dana i ima relativno malo mogućih odstupanja. S druge strane, danas pažnju izazivaju uglavnom sistemi koji uče iz podataka, kod kojih ne postoje eksplicitna uputstva za ponašanje. Umesto direktno saopštenih striktnih pravila, AI sistemi koriste velike količine podataka da nađu pravilnosti koje mogu da iskoriste da savladaju zadatak, što je sličnije načinu na koji ljudi uče. Na primer, ljudski vozač nakon određenog broja sati vožnje prepoznaje kako da reaguje u različitim situacijama: recimo, uočio je pešaka na prelazu i procenio da treba da uspori. Slično, AI sistemi u autonomnom, samovozećem automobilu koriste senzore umesto očiju i na osnovu snimaka ljudske vožnje znaju koja je prikladna akcija na dati scenario. Prednost ovakvog pristupa, zasnovanog na učenju iz podataka, jeste što nije neophodno poimence pobrojati sve moguće scenarije i željene reakcije sistema u jednoj promenljivoj i kompleksnoj ljudskoj radnji kao što je vožnja.

Već ovde možemo videti značaj podataka koji se koriste za obuku ovakvih AI sistema. Nakon definisanja cilja AI sistema, prvi korak u njegovom razvoju jeste prikupljanje odgovarajućih podataka za učenje. (U većini slučajeva, proces učenja podrazumeva određivanje parametara matematičkih modela koji donose odluke u novim situacijama na osnovu prethodno viđenih podataka.) Baza podataka na osnovu koje se AI sistem obučava naziva se *skup za obuku* (engl. *training dataset*). Zavisno od načina primene, potrebno je labelirati podatke, odnosno označiti ih na način koji je pogodan za mašinu. Na primeru vožnje, ovo bi bili video-snimci ulice na kojoj su označeni važni elementi: put, ivičnjak, semafor, drugi automobili, pešaci, prepreke (Slika 1).

Slika 1. Primer isečka iz videa u kom su obeleženi relevantni objekti (put, vozilo).
https://raw.githubusercontent.com/deetungsten/website-images/master/ml/segmanticexample.png

Zavisno od prirode problema, ulazni podaci mogu biti raznovrsni: slika, video, tekst, zvuk… Naravno, moguće je i kombinovati različite ulaze. Značaj pažljivog izbora i obeležavanja podataka se oslikava u principu poznatom kao *smeće na ulazu, smeće na izlazu* (engl. *garbage in, garbage out*). U primeru sa autonomnim automobilom, ukoliko mašina "vidi" previše primera rizične vožnje ili nedovoljno obeleženih pešaka, ne može se očekivati da krajnji ishod učenja mašine u vožnji bude dobar. Više o tipovima podataka i njihovoj pravilnoj upotrebi pročitajte u drugom i četvrtom poglavlju. Nakon što je obuka modela završena, potrebno je proveriti uspešnost AI sistema, kroz proces *testiranja*. U ovoj fazi se proverava koliko je obučeni model sposoban za generalizaciju tako što se izlaže novim podacima, iz skupa za testiranje, koji nisu korišćeni tokom obuke.

ZAŠTO JE VEŠTAČKA INTELIGENCIJA VAŽNA?

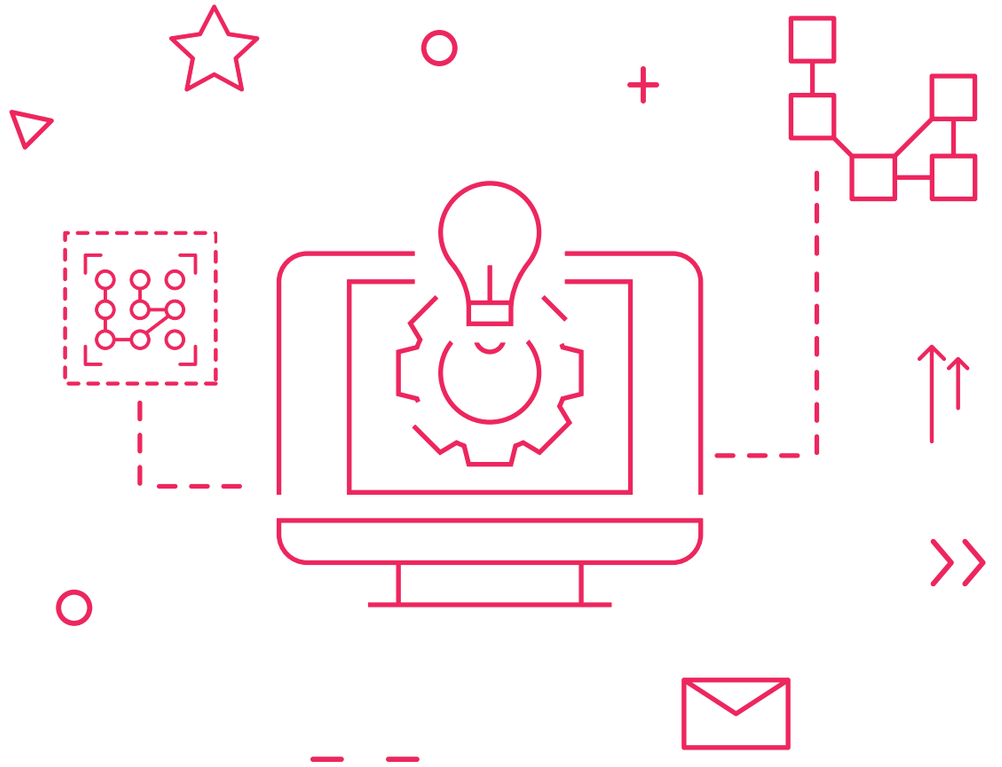
Već danas, prosečan stanovnik Zemlje je verovatno izložen AI sistemima nekoliko puta tokom jednog dana. Jutarnje vesti i večernju Netfliks seriju mu predlažu sistemi preporuke, koji funkcionišu na osnovu informacija o sadržajima koji su ga prethodno interesovali i prosečnog profila korisnika. Čak i ako ne živi u zemlji gde je dostupno delimično autonomno Tesla vozilo, moguće da mu pri parkiranju pomaže ugrađeni sistem koji na osnovu računrskog vida (engl. *computer vision*) prepoznaje prepreke i meri rastojanje do istih. Ukoliko želi da promeni posao, moguće je da će njegov CV prvo analizirati mašina, upotrebom tehnika obrade prirodnog teksta (engl. *natural language processing*). Ne pronalazite se u ovim primerima? Ako ste guglali bilo šta u toku dana, izvesno je da ste bili indirektni korisnik AI sistema.

Trenutno se nalazimo u fazi tzv. Četvrte industrijske revolucije, u kojoj značajnu ulogu igra upravo veštačka inteligencija. Industrijske revolucije su periodi tokom kojih usled inovacija dolazi do nagle promene korišćenih alata i posledično promene na tržištu rada i u svakodnevnom životu. Broj oblasti u kojima se primenjuje veštačka inteligencija se poslednjih godina strmoglavo povećava, od automatizacije procesa u industriji i široke primene u IT sektoru, preko nauke i medicine do ekonomije i prava (Slika 2).

Slika 2. Primeri oblasti u kojima se koristi veštačka inteligencija.

Iz ugla srpskog IT sektora, AI predstavlja šansu kako za startape, tako i za već ustanovljena preduzeća. Nulti preduslov je postojanje odgovarajuće obrazovanih stručnjaka iz ove oblasti, što je delimično rešeno uvođenjem odgovarajućih fakultetskih predmeta i master studija iz AI. Drugo ograničenje je specijalizovana infrastruktura koja je neretko potrebna za obradu velike količne podataka koja se koristi u AI. Sa razvojem tzv. računarstva u oblaku (engl. *cloud computing*) više nije neophodna velika početna investicija za infrastrukturu koja podržava AI modele kako bi se pariralo konkurenciji iz razvijenijih zemalja. Više o tome kako izgleda tok razvoja jednog AI sistema možete pročitati u trećem poglavlju.

Zašto je za prosečnog građanina važno razumevanje ideja i primena AI? Pored opšte pismenosti u vreme kada tehnologija postaje neodvojivi deo sadašnjosti, sve više ćemo se susretati sa AI i u profesionalnom kontekstu, čak i u zanimanjima naizgled daleko od IT sektora. Kao što za druge tehnološke discipline postoje jasno definisani legalni okviri - poput građevinskih propisa ili uredbi o kvalitetu namirnica - tako trenutno postaje sve jasnije da i za AI mora da postoji određena forma revizije kako bi se osigurala zaštita korisnika i njihovih podataka. (Više o ovom pitanju možete pročitati u poglavlju o etici.)



MITOVI I ČINJENICE U VEZI SA VESTAČKOM INTELIGENCIJOM

MIT: Veštačka inteligencija je na putu da prevaziđe čovekovu (preterani optimizam).

ČINJENICA: Trenutno smo daleko od veštačke inteligencije koja je uporediva sa čovekovom. U brojnim specijalizovanim, jasno definisanim zadacima AI može da bude uspešniji od čoveka - na primer, u šahu, pokeru, pa čak i pre-poznavanju slika, tumora i kompleksnoj onlajn igri Dota. Ipak, mogućnost integracije različitih vrsta inteligencije zasad ostaje van domašaja veštačke inteligencije.

MIT: AI može da bude koristan samo za repetitivne poslove (preterani pesimizam).

ČINJENICA: Jedan moguć kontraprimer su govorni asistenti, koji odgovaraju na pitanja i sposobni su da održe osnovnu konverzaciju. Štaviše, do sada se pokazalo da čak i određeni broj kreativnih poslova, poput crtanja i pisanja, može da se reprodukuje pomoću AI rešenja.

MIT: Veštačka inteligencija i mašinsko učenje su isto.

ČINJENICA: Iako ne postoji precizna definicija ovih oblasti jer se brzo menjaju, u pitanju su dva različita pojma. Veštačka inteligencija obuhvata širi skup koji obuhvata sve sisteme koji oponašaju prirodnu inteligenciju, dok se mašinsko učenje odnosi na oblast u kojoj se koriste podaci za učenje pravilnosti.

MIT: Mašine uče same od sebe.

ČINJENICA: Pre svega, cilj AI sistema postavlja tim koji ga projektuje, mašine nemaju svest. Takođe, za razvoj sistema koji koristi veštačku inteligenciju su potrebni ulazni podaci i algoritmi, koje takođe određuju ljudi koji projektuju ove sisteme.

MIT: AI su roboti.

ČINJENICA: AI su i antropomorfni roboti poput C3PO iz “Ratova zvezda”, ali i mnogo šta što nije prva asocijacija: softver za prepoznavanje objekata na slikama, botovi koji proveravaju da li je pristigli mejl spam, kao i pomenuti autonomni automobili i govorni asistenti.

OSNOVNA AI TERMINOLOGIJA

AI

Veštačka inteligencija

BOT

Autonomni agent sposoban za interakciju (npr. kroz formu četa).

MAŠINSKO UČENJE *MACHINE LEARNING*

Podskup AI i oblast računarstva koja se bavi učenjem iz podataka bez eksplicitno zadatih instrukcija.

NEURONSKA MREŽA *NEURAL NETWORK*

Osnovni koncept modela neuronskih mreža je zasnovan na ideji bioloških neurona povezanih sinapsama. Osnovna jedinica računanja je jedan (veštački) neuron. Svaka neuronska mreža ima ulazni i izlazni sloj koji čine proizvoljan broj neurona, a između se nalazi tzv. skriveni slojevi. Tzv. plitke neuronske mreže imaju tačno jedan skriveni sloj, dok duboke neuronske mreže imaju bar dva, a obično daleko više slojeva.

OBUKA

Proces tokom kog se na osnovu podataka i pomoću zadatog algoritma određuju parametri modela.

DUBOKO UČENJE *DEEP LEARNING*

Oblast u kojoj se koriste duboke neuralne mreže (engl. deep neural networks) kao modeli (videti i: neuronske mreže). Zavisno od načina primene i tipa ulaznih podataka, neke od najčešće korišćenih tipova dubokih neuralnih mreža su konvolucione (engl. CNN), rekurentne (engl. RNN), mreže sa dugom kratkotrajnom memorijom (engl. LSTM), generativne suparničke mreže (engl. GAN), itd.

PARAMETRI MODELA

Skup brojeva koji opisuju model. U jednostavnom primeru linearne zavisnosti, postoje samo dva parametra modela: nagib i odsečak prave. U slučaju dubokih neuronskih mreža, broj parametara može biti reda veličine nekoliko miliona.

SINGULARITET

Hipotetički trenutak u kom mašine prestižu ljudsku inteligenciju i samim tim postaju samodovoljne. Pojam je definisao Džon fon Nojman pedesetih godina XX veka, međutim ne postoji pouzdana procena da li će, i kada, doći do takvog ishoda.

VAŽNOST
PODATAKA

ISTORIJSKI OSVRT

Problem skladištenja, organizacije i korišćenja podataka su izazovi sa kojim su se ljudi sretali mnogo pre modernog sveta, pa se prvi primeri o konkretnim rešenjima mogu naći na čuvenim kostima iz Išanga, koje su, ukoliko je pretpostavka istraživača o njihovoj nameni dobra, primer najranijih skladišta podataka. Pretpostavka je da su plemenski ljudi iz Išanga pre oko 20.000 godina urezivali zareze na kostima kako bi vodili evidenciju o trgovinskim aktivnostima, ali i kako bi evidentirali zalihe hrane i planirali njeno korišćenje, odnosno predviđali koliko će te zalihe trajati.

Slika 2.1: Kosti iz Išanga

Bila ona u potpunosti istinita ili ne, priča o kostima iz Išanga dobro oslikava suštinu - smislene podatke je korisno čuvati i koristiti ih u cilju poboljšavanja kvaliteta procesa koji su određeni tim podacima. Dalje, od abakusa, koji je prvi zabeleženi primer pomagala za računanje, preko stvaranja biblioteka (aleksandrijska biblioteka i ostale tog, ali i kasnijeg doba), koje su dobar primer potrebe ljudi da skladište podatke na neki standardizovan i dobro uređeni način, radi bržeg pristupa tim podacima kasnije; pa sve do primera sa kraja 19. veka kada je bilo potrebno obraditi ogromne količine podataka prikupljenih prilikom popisa stanovništva u SAD. Obrada podataka sa popisa iz 1880. godine trajala je 7 godina. S obzirom na porast stanovništva u SAD 1890. procena je bila da bi obrada tih novih podataka trajala i nakon sledećeg popisa zakazanog za 1900. godinu, što nije bilo prihvatljivo, te je bilo potrebno osmisliti novo rešenje. Tada je Herman Holerit razvio sistem, koji se sastojao od mašinski čitljivog medijuma - bušenih kartica i mašine koja je te podatke mogla brzo (u kontekstu tog perioda) obraditi. Ovaj sistem je uspešno obradio podatke za godinu dana (na temeljima tog sistema i Holeritove male kompanije, nastala je čuvena kompanija IBM).

Jedna od najznačnijih ličnosti za oblast računarstva i informatike, ali i za postavljanje osnova veštačke inteligencije, je Alan Tjuring, široj javnosti najpoznatiji po svojoj zasluzi za dešifrovanje “Enigme”, mašine koju je nemačka armija koristila u Drugom svetskom ratu za sigurno slanje šifrovanih poruka. Tjuring je, između ostalih dostignuća, 1950. godine osmislio test koji je trebalo da odgovori na pitanje da li mašine mogu razmišljati poput ljudi. Danas ovaj test znamo kao Tjuringov test. Za ovaj test je potrebno tri učesnika: 2 osobe i računar. Jedna osoba je ispitivač i ona upućuje pitanja osobi i računaru, a sa ciljem da otkrije koji učesnik testa je osoba, a ko je računar. Komunikacija se odvija putem računara, dakle slanjem i odgovaranjem na poruke. Računaru je dozvoljeno da radi sve kako bi prevario ispitivača, na primer na pitanje “Da li si računar?” može odgovoriti sa “Ne”, a nakon toga i netačno (ili sporo; ili i

sporo i pogrešno) odgovoriti na neko matematičko pitanje, poput množenja velikih brojeva, na koje bi podrazumevano računar mogao da odgovori i brzo i tačno. Test se ponavlja određeni broj puta i ako osobe u ulozi ispitivača greše dovoljno često u odabiru ko je čovek, a ko mašina, računar se smatra entitetom sposobnim za razmišljanje. Do danas, nijedan računar nije "položio" Tjuringov test.

O veštačkoj inteligenciji je prvi put diskutovano kao o samostalnoj informatičkoj disciplini na konferenciji u Dartmutu (SAD) 1956. godine. Tada je i skovano ime za ovu oblast na predlog Džona Makartija. U početku su metode veštačke inteligencije uglavnom bile zasnovane na deduktivnom zaključivanju - u ovom pristupu algoritmi za rešavanje problema su eksplicitno opisani i moguće ih je strogo matematički analizirati (i svaki pojedinačni primer jednog problema je moguće analizirati na taj način, npr. najkraći put od tačke A do tačke B se može analitički odrediti). U ovu grupu algoritama spadaju, primera radi, algoritmi pretrage, genetski algoritmi, automatsko rasuđivanje, itd (više o ovim metodama i principima se može pročitati u knjizi "Veštačka Inteligencija", navedenoj u dodatnim resursima). Danas su postale dominantne metode zasnovane na induktivnom zaključivanju, koje je "vođeno" podacima i nije uvek moguće eksplicitno opisati proces rešavanja pojedinačnih primera jednog problema. Uzmimo za primer problem klasifikacije objekata - na slici se nalazi roze lopta. Kod induktivnog zaključivanja, nije moguće sa sigurnošću tvrditi da je obučeni model klasifikovao konkretnu roze loptu kao loptu zato što je *lopta okrugla* ili zato što je *lopta roze*, pošto sve zavisi od toga nad kojim podacima je model učio šta je lopta. Drugim rečima, ukoliko su sve lopte koje je model "video" bile roze boje, ne možemo sa sigurnošću tvrditi da će dobro klasifikovati i lopte drugih boja. Razvoj ovog tipa metoda (mašinsko učenje, duboko učenje, neuronske mreže) doveo je do ostvarivanja odličnih rezultata u raznim oblastima poput računarskog vida, automatskog upravljanja vozilima, automatskog prevođenja, itd. Dominaciju drugog pristupa omogućili su sve moćniji procesori, koji su vremenom i prilagođavani tako da efikasno izvršavaju ključne matematičke operacije za proces "učenja" algoritama (*CPU* << *GPU* << *TPU*, više o svakom tipu procesora, kao i njihovo poređenje se može pročitati u Guglovom blogu navedenom u dodatnim resursima). Pretpostavlja se da će se u budućnosti ova dva pristupa integrisati, odnosno da će sistemi veštačke inteligencije kreirati modele, koji će pritom moći eksplicitno da opišu postupak rešavanja pojedinačnih slučajeva (odnosno dobićemo odgovor na pitanje *zašto je roze lopta klasifikovana kao lopta*).

Kada se osvrnemo na sve pomenuto, možemo primetiti da je rad sa podacima uvek bio prisutan, dok su se mogućnosti i granice neprestano pomerale. Sa razvojem nauke i pre svega, tehnologije, koja bi podržala zahtevne matematičke operacije i skladišne potrebe, dolazi do procvata oblasti veštačke inteligencije, pre svega zasnovanoj na induktivnom zaključivanju, čemu svedočimo danas.

RAZLIČITI TIPOVI PODATAKA

Neformalno govoreći, skup podataka se može sastojati od različitih vrsta objekata poput tekstova, slika, video ili audio zapisa, slogova u bazi, vremenskih serija, itd. Skupovi podataka se dakle međusobno mogu razlikovati na mnogo načina, odnosno postoji veliki broj različitih tipova podataka. Formalnije, skupovi podataka (engl. *datasets*) se mogu posmatrati kao skupovi objekata (uzoraka) koji imaju određene atribute (engl. *features*) koji bliže opisuju posmatrane objekte, odnosno njihove karakteristike.

ATRIBUTI

Atributi (obeležja) predstavljaju karakteristike objekata, koje mogu varirati u zavisnosti od posmatranog objekta (različite osobe imaju npr. različit atribut "datum rođenja") ili u zavisnosti od vremena (ista osoba u različitim periodima života, npr. atribut "telesna temperatura"). Jedan način na koji možemo razlikovati atribute jeste po tome koje se operacije mogu primeniti na njima:

RAZLIČITOST

Da li su dva objekta na osnovu posmatranog atributa ista ili različita.

POREĐENJE

Koji objekat je veći odnosno manji na osnovu posmatranog atributa.

ADITIVNOST

Da li atribut podržava operacije sabiranja/oduzimanja.

MULTIPLIKATIVNOST

Da li atribut podržava operacije množenja/deljenja.

Možemo definisati četiri tipa atributa na osnovu toga koje od ovih operacija "podržavaju":

IMENSKI *NOMINAL*

Imenski ili kategorički atribut podrazumeva mogućnost razlikovanja jednog objekta od drugog (podržana operacija: "različitost"). Primer imenskog atributa može biti međunarodni standardni knjižni broj (ISBN) pomoću kojeg možemo zaključiti da li su neke dve knjige na osnovu ovog atributa iste ili različite, dok dodavanje ili oduzimanje nekog broja od tog broja nije moguće, niti bi imalo nikakvog smisla. Još neki primeri imenskih atributa su poštanski brojevi, ali i boja očiju, pol, zanimanje, mesto rođenja, i slično.

REDNI *ORDINAL*

Atribut ovog tipa nam daje dovoljno informacija kako bismo objekte mogli urediti (podržane operacije: "različitost", "uređenje"). Restorane na online servisima za preporuku možemo urediti po rejtingu, pa je ocena primer jednog rednog atributa. Takođe, primer rednog atributa je i starosna kategorija u sportskim takmičenjima (petlići, pioniri, kadeti) - objekte (u ovom slučaju ljude) možemo urediti po kategorijama od najmlađih ka starijim (ili obrnuto).

INTERVALNI *INTERVAL*

Intervalne atribute možemo posmatrati kao vrednosti koje nam omogućavaju da odredimo koliko su neki objekti različiti (podržane operacije: “različitost”, “uređenje”, “aditivnost”). Na primer, znamo da od 25. juna do 2. jula prođe 7 dana, pa je upravo datum jedan primer intervalnog atributa. Još jedan primer intervalnog atributa je temperatura u Celzijusima ili Farenhajtima.

RAZMERNI *RATIO*

Kod razmernih atributa moguće je primenjivati sve navedene operacije - “različitost”, “uređenje”, “aditivnost”, “multiplikativnost”. Samim tim, moguće je utvrditi odnos objekata, npr. koliko puta je jedna vrednost veća od druge. Primeri razmernih atributa su temperatura u Kelvinima, godine, masa, broj listova na drvetu, itd.

Imenske i redne atribute možemo posmatrati kao kategoričke (kvalitativne atribute) dok intervalne i razmerne atribute možemo posmatrati kao numeričke (kvantitativne) atribute. Atributi se mogu podeliti i po tome koliki im je skup vrednosti koje mogu imati. Dve grupe atributa po ovoj podeli su:

DISKRETNI ATRIBUTI

Imaju konačan skup vrednosti. Uglavnom su ovde u pitanju celobrojne promenljive, npr. poštanski brojevi, broj stana, ali je primer diskretnih atributa i skup reči ispisanih u nekoj knjizi s obzirom na to da je skup tih reči konačan (možemo zamisliti da svaku reč iz te knjige možemo predstaviti tačno jednim brojem, pa i ovakav skup možemo svesti na celobrojne vrednosti). Takođe, binarni atributi predstavljaju specijalan slučaj diskretnih atributa kod kojih postoje samo dve vrednosti (stanje prekidača, ishod teniskog meča, itd).

NEPREKIDNI ATRIBUTI

Skup vrednosti odgovara skupu realnih brojeva, te su primeri ovakvih atributa: temperatura, brzina, visina, itd.

SKUPOVI PODATAKA

Kao što je već napomenuto, postoji veliki broj tipova podataka, a samim tim i skupova podataka. Ipak, sve tipove je moguće grupisati, pa ćemo u nastavku govoriti o tri grupe podataka, iako ovo nije ni jedini mogući način grupisanja niti su ovo jedine moguće smislene grupe: tabelarni podaci, grafovski podaci (podaci zasnovani na grafovima), podaci sa poretkom.

TABELARNI PODACI

Jedan zapis (red) u tabeli predstavlja objekat koji ima određene karakteristike predstavljene njegovim atributima (kolone). Tabelarni podaci se sastoje od skupa zapisa odnosno skupa objekata, pri čemu je svaki objekat opisan istim skupom atributa (svaki pasoš ima unapred definisane atribute). Ovakvi podaci se najčešće skladište u relacionim bazama podataka ili u datotekama, danas sve češće u datotekama koje sadrže vrednosti odvojene zarezom (engl. *Comma-separated Values* - CSV), koje imaju razne pogodnosti u odnosu na relacione baze podataka - jednostavna čitljivost i mogućnost izmena, brže generisanje jednostavnog dokumenta, može se procesirati korišćenjem bilo kojeg alata, itd.

ID	DOMAĆIN	GOST	GOLOVI DOMAĆINA	GOLOVI GOSTA	DATUM	POSETA	SUDIJA
111	LIVERPOOL	MAN UNITED	2	0	31/03/01	44,806	GRAHAM POLL
112	MAN CITY	ASTON VILLA	1	3	31/03/01	34,240	ROB STYLES
113	SUNDERLAND	LEEDS	0	2	31/03/01	48,285	STEVE DUNN
114	WEST HAM	EVERTON	0	2	31/03/01	26,044	ANDY D' URSO
115	COVENTRY	DERBY	2	0	31/03/01	19,654	DAVID ELLERAY

Tabela 2.1: Primer tabelarnih podataka

Transakcioni podaci su specijalna vrsta tabelarnih podataka za koje važi da svaki objekat, odnosno svaka transakcija sadrži određeni skup stavki. Potrošačka korpa je jednostavan primer transakcionih podataka pri čemu se na svakom računu (transakciji) mogu naći različite stavke.

ID TRANSAKCIJE	STAVKE
2526	HLEB, MLEKO, SO
2527	ŠEĆER, ULJE, BRAŠNO
2528	PIVO, HLEB, PELENE, KISELA VODA
2529	KISELA VODA, RAKIJA

Tabela 2.2: Primer transakcionih podataka - potrošačka korpa

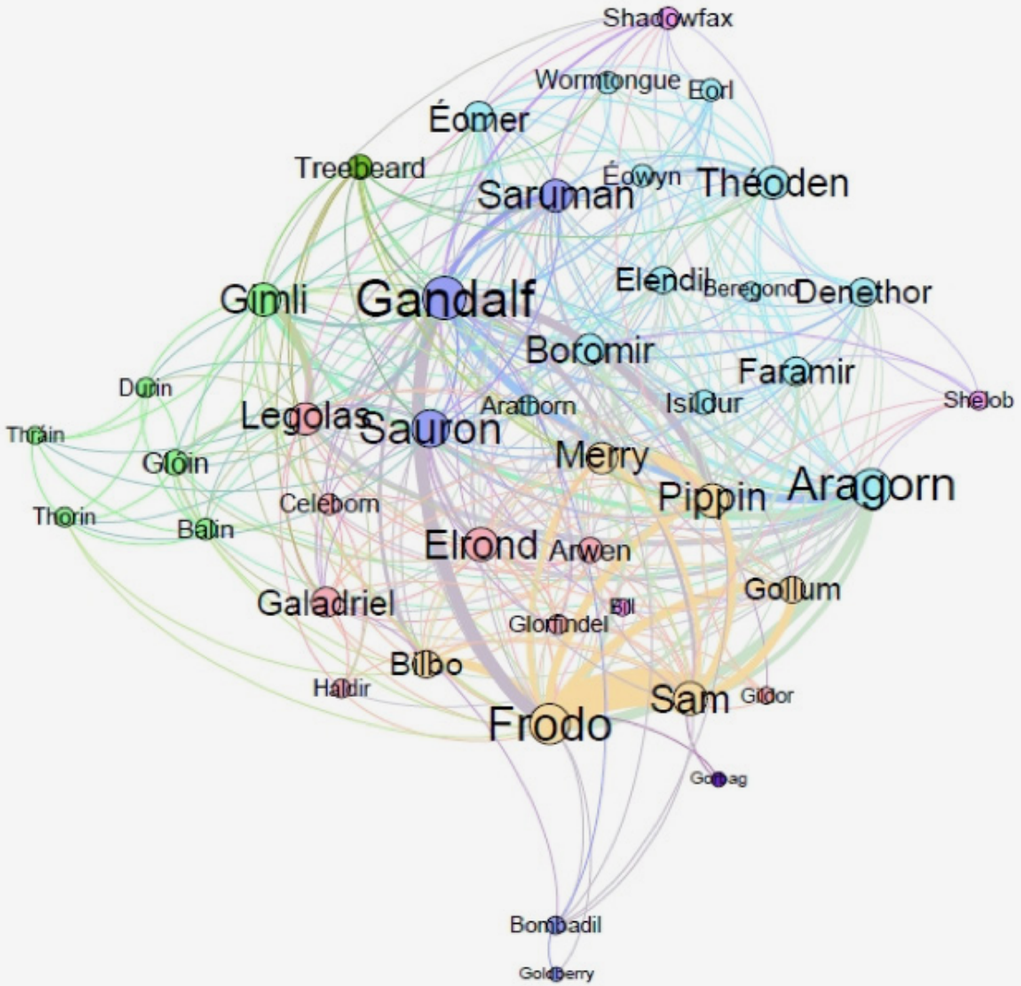
Još jedan primer tabelarnih podataka je slučaj kada svi objekti imaju fiksirani skup numeričkih atributa. Tada, objekte možemo posmatrati kao tačke u višedimenzionalnom prostoru ili vektore, pri čemu svaki atribut opisuje jednu dimenziju. Ovakav tip podataka nazivamo *matrica podataka*. Kao što ime kaže, ovakav skup podataka možemo predstaviti i pomoću *m x n* matrice gde m predstavlja broj objekata, a *n* broj atributa. Uzimajući u obzir ove karakteristike, nad ovakvim podacima je moguće izvršavati matrične operacije, te je ovakav format često vrlo pogodan za predstavljanje statističkih podataka.

GEO. ŠIRINA	GEO. DUŽINA	TEŽINA (KG)	PRIORITET
44.7835985	20.4614069	750,35	4
44.8164305	20.450322	254,12	1
44.8277205	20.4041697	322,01	3
44.8041182	20.4894575	412,75	1

Tabela 2.3: Primer matrice podataka (lokacije sa prijavljenim đubretom koje je potrebno pokupiti korišćenjem zamišljene aplikacije za prijavu i skupljanje đubreta)

GRAFOVSKI PODACI

Graf je struktura podataka, koja se sastoji od čvorova i grana, pri čemu su objekti predstavljeni čvorovima, a granama su predstavljeni odnosi između tih objekata. Grane mogu biti usmerene ili neusmerene i sa određenim težinama. Ovakva struktura često može biti pogodna za reprezentaciju podataka. Saobraćajnice mogu biti jedan primer grafovskih podataka, gde su raskrsnice čvorovi, a ulice grane (jednosmerne ulice - usmerene grane, dvosmerne ulice - neusmerene grane), dok bi težina mogla biti dužina ulice, ali bi se u računanje težine ulice mogle dodati i razne druge karakteristike (broj traka, broj semafora, ograničenja, itd.).



Slika 2.2: Primer grafovskih podataka - Odnosi likova iz romana “Gospodar prstenova” izvučeni na osnovu njihove međusobne interakcije i pojavljivanja u bliskim delovima teksta

Primer grafovskih podataka koje prosečan korisnik interneta najčešće koristi (implicitno ili eksplicitno odnosno svesno ili nesvesno) su društvene mreže gde pojedinačni korisnici mogu biti (i često jesu) predstavljeni čvorovima, a njihovi odnosi granama koje predstavljaju određeni vid interakcije. Mašinsko učenje koje radi sa grafovskim podacima - *grafovsko mašinsko učenje*, je prethodnih godina doživelo pravi procvat i sve je više realnih primena ovog principa u oblastima poput saobraćaja, hemije, verifikacije softvera, sistema za preporuku.

PODACI SA PORETKOM

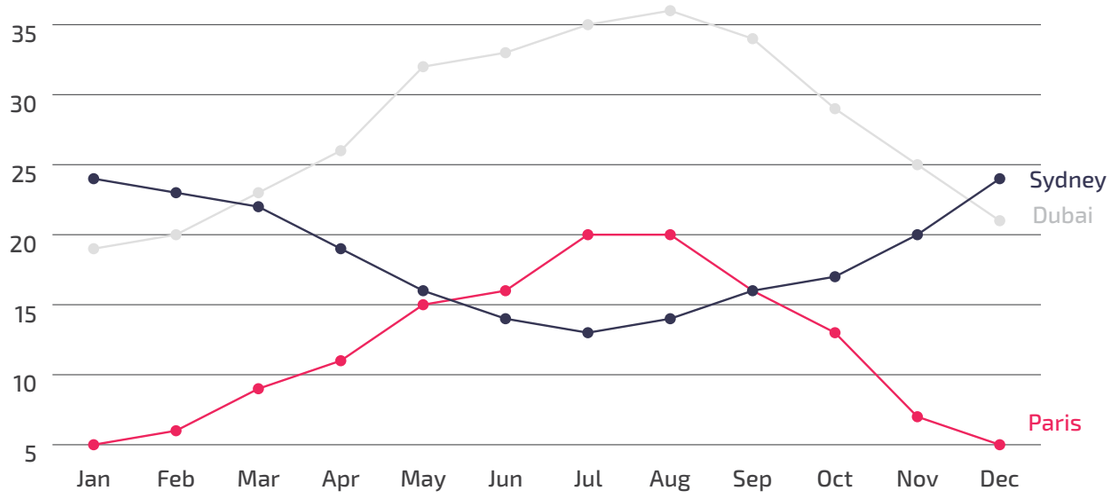
Odnos atributa ove grupe podataka je predstavljen u nekom vremenskom ili prostornom poretku. Ukoliko na primer gore opisanim transakcionim podacima dodamo atribut *vreme kupovine*, dobijamo podatke sa poretkom, koji su time "dobili" vremensko uređenje. Ovakav tip podataka nazivamo redni (engl. sequential) ili vremenski podaci (engl. *temporal data*). Primera radi, koristeći atribut vreme kupovine možemo izvlačiti zaključke o tome koji artikli su popularni u koje doba godine i shodno tome pripremati zalihe.

Sledeći primer podataka sa poretkom su podaci u vidu nizova, sekvencijalni podaci (engl. *sequence data*). Njih karakteriše niz (sekvenca) individualnih objekata poput reči, slova i slično. Iako idejno podsećaju na gore opisane redne podatke, u ovom slučaju ne postoji uređeni vremenski redosled, već svaki objekat zauzima svoju poziciju u nizu. Primer ovog tipa je reprezentacija gena u vidu alternirajućeg niza četiri različita nukleotida od kojih je sastavljena DNK (Adenin, Timin, Guanin i Citozin).

GGGGGAGGCGGGGCCGCCCGAGC
CCAACCGAGTCCGACCAGGTGCC
CCCTCTGCTCGGCCTAGACCTGA

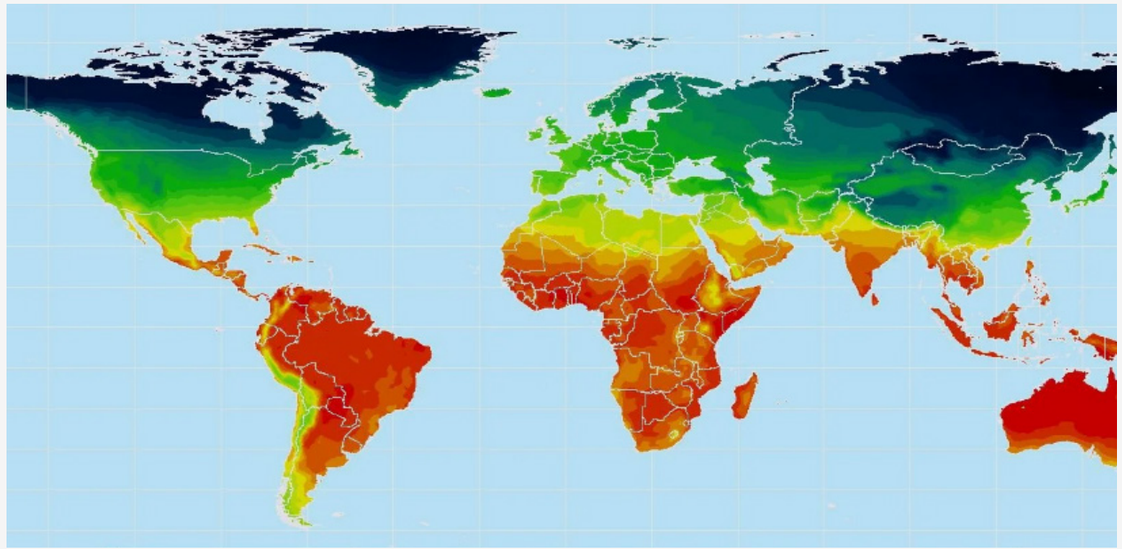
Slika 2.3: Podaci u vidu nizova - primer sekvence nukleotida

Vremenski podaci (engl. *time series data*) predstavljaju tip podataka sa poretkom gde je svaki objekat vremenska serija, odnosno niz merenja, koja se odvijaju u različitim trenucima. Primer vremenskih serija je skup podataka, koji predstavlja količinu padavina kroz dane, mesece, godine ili broj sunčanih dana, visina plime i oseke kroz određene vremenske periode. Vremenske serije se koriste u statistici, obradi signala, astronomiji, seizmologiji, itd.



Slika 2.4: Primer vremenske serije - Temperatura po mesecima na tri različite lokacije

Prostorni podaci (engl. *spatial data*) su oni podaci koji imaju prostorne atribute, poput lokacija na mapi ili određenih površina. Primer prostornih podataka su podaci o vremenskim prilikama - padavine, temperatura, vazdušni pritisak. Jedna od osobina prostornih podataka je prostorna autokorelacija - objekti koji su blizu u prostoru imaju određene sličnosti (npr. bliska mesta imaju slične vremenske uslove).



Slika 2.5: Primer prostornih podataka - srednja temperatura u januaru na raznim lokacijama

KVALITET PODATAKA

Podaci najčešće dolaze sa nekom greškom, bilo da je ta greška nastala prilikom merenja (merenje senzorima ili jednostavno slučajni ili namerni previdi ljudi prilikom popunjavanja anketa za sprovođenje određenih socijalnih istraživanja) ili je celokupan problem loše definisan ili loše organizovan, pa se na kraju procesa prikupljanja podataka završi sa gomilom neupotrebljivih ili sa delimično upotrebljivim podacima. Iako se određene greške mogu tolerisati ili kasnije otkloniti na neki način, potrebno je veliku pažnju posvetiti kvalitetu podataka nad kojima će se razvijati modeli, jer će upravo ti podaci činiti njihovu osnovu - iz loših podataka nije moguće izvesti korisne zaključke. Štaviše, može se upasti u zamku i koristiti model, koji je razvijen nad lošim podacima, jer se "čini" da model daje korektne rezultate. Proces prikupljanja podataka je potrebno dobro isplanirati i organizovati, dok je prilikom korišćenja "tuđih" podataka potrebno biti oprezan, a u svakom slučaju, neophodno je dobro pretprocesirati podatke (prečistiti ih i učiniti pogodnim za dalju analizu). Sa druge strane, treba napomenuti da dobri podaci ne garantuju dobar model.

SKLADIŠTENJE PODATAKA

U kontekstu skladištenja podataka, podatke možemo grubo podeliti u dve kategorije - struktuirani i nestruktuirani podaci. Suština struktuiranih podataka je da su dobro organizovani i može im se brzo pristupiti (npr. podaci skladišteni u relacionim bazama podataka), dok su nestruktuirani podaci neorganizovani, odnosno njihovo skladištenje ne prati neku unapred definisanu šemu (tekstualni dokumenti, slike i ostali multimedijalni sadržaj, mejlovi, itd). Vredi napomenuti da to što su podaci struktuirani ne znači da su istovremeno i smisleni. Sa druge strane, nestruktuirani podaci ne moraju nužno biti "potpuno razbacani na sve strane", moguća je osnovna organizacija na

osnovu metapodataka (“podaci o podacima” - opisuju karakteristike datoteka, engl. *metadata)* radi lakše pretrage i korišćenja kasnije. Može se reći da su to polustruktuirani podaci.

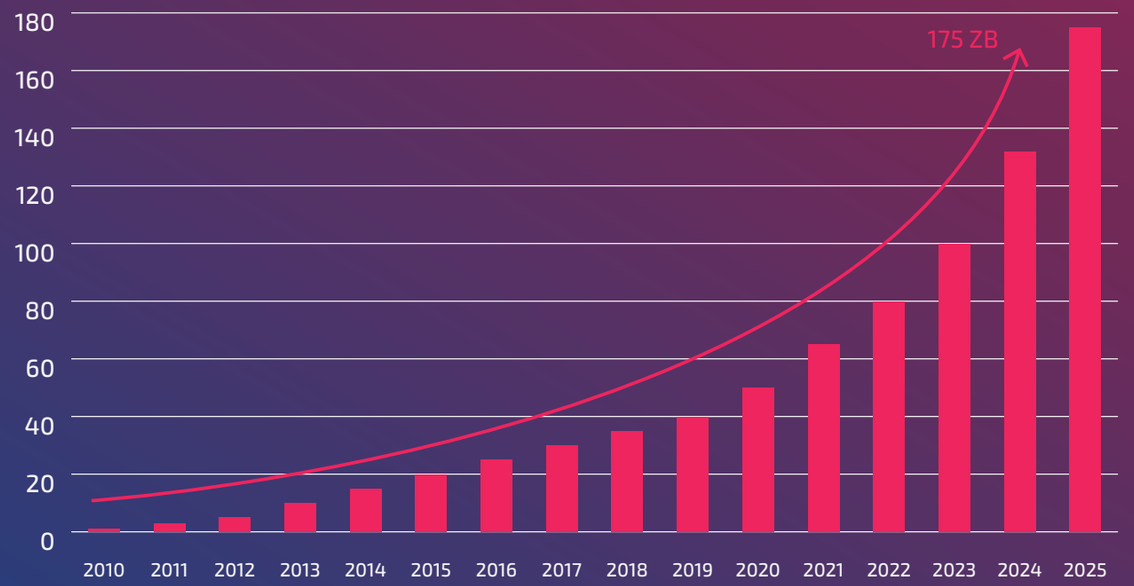
U prošlosti, kada je čuvanje podataka bilo skupo ($500.000 po gigabajtu 1980. godine) praksa je bila da se radi minimizacije troškova uglavnom pribegava čuvanju struktuiranih podataka, odnosno da se “višak” informacija odbacuje. Vremenom, kako su troškovi za skladištenje podataka postajali sve manji, čuvano je sve više nestruktuiranih podataka. Danas, kada je čuvanje jednog terabajta podataka oko 10 miliona puta jeftinije nego osamdesetih i s obzirom na to da je uz razvoj metoda veštačke inteligencije olakšana analiza takvih podataka, baš ono što se u početku smatralo “viškom” sve češće je ključ za transformisanje poslova, poslovnih odluka i postizanje većih poslovnih uspeha. Što se tiče konkretnih situacija i potreba oko skladištenje podataka u praksi, one su vrlo različite u zavisnosti od toga:

Ko izvršava projekat - Ukoliko je u pitanju samostalni istraživački projekat, uglavnom je dovoljno podatke čuvati na laptopu gde se projekat i razvija, dok je kod malih ili velikih preduzeća uglavnom potrebno mnogo više prostora, pa se podaci čuvaju na većim diskovima (lokalno ili na udaljenim serverima);

Koja je faza projekta - Najčešće nije neophodno u početnoj fazi projekta pribegavati skupim rešenjima, već je projekat moguće razvijati iterativno (korak po korak) i skladišta i korišćene resurse uvećavati u skladu sa potrebama.

Kako bi se stekla slika o tome koliko novih podataka se generiše svake godine i koliki izazov predstavlja skladištenje podataka danas, dovoljno je navesti dve činjenice:

- **90% globalnih podataka je nastalo u prethodne 2 godine;**
- **Otprilike na svake 2 godine se količina globalnih podataka duplira.**

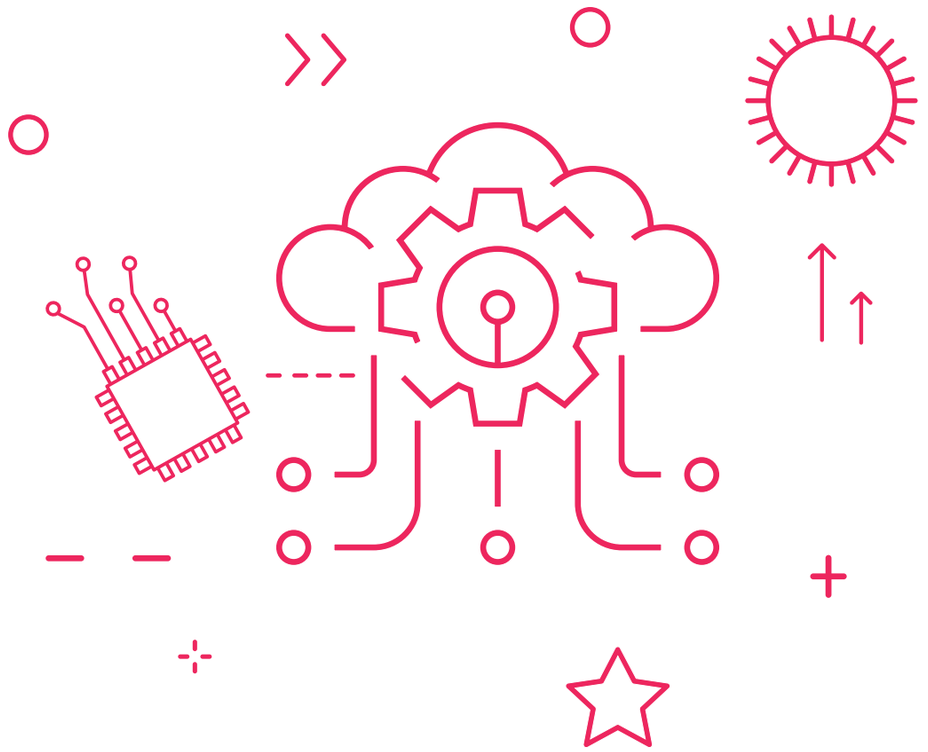


Slika 2.6: Porast količine podataka na globalnom nivou i predviđeni rast do 2025. godine (u zetabajtima, 1 zetabajt = 1.073.741.824 terabajta)

Rastuće količine podataka (pretežno nestruktuiranih) u skoro svim relevantnim oblastima, dovele su do toga da se jedna fraza sve češće čuje - *Big Data*. Ova fraza označaca skupove podataka koji su izuzetno obimni i pritom neprestano rastu velikom brzinom. Primer ovakvog skupa podataka je njujorška berza, koja generiše približno terabajt novih podataka dnevno. Takođe, motor prosečne komercijalne letelice, poput aviona Boing 737, generiše 20 terabajta podataka po satu leta.

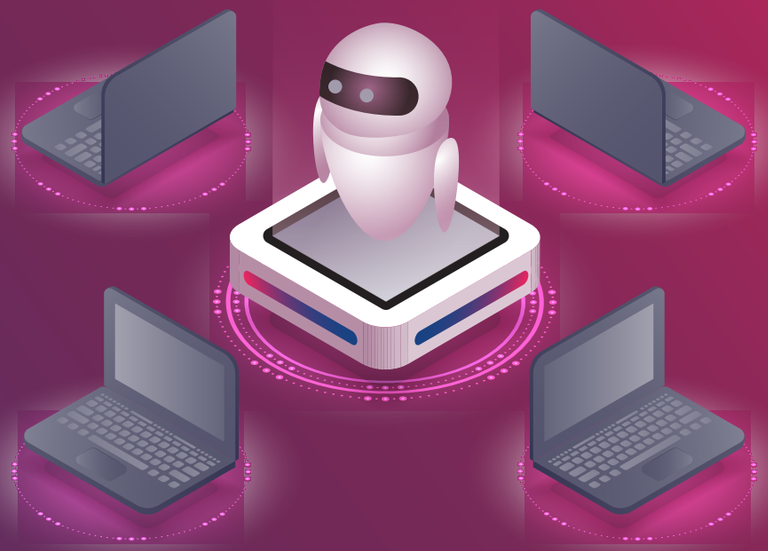
Postoji više tipova skladišta podataka u zavisnosti od tipa i količine podataka, koje je potrebno čuvati. Neki od tipova skladišta su: baza podataka (eng. *database*), skladište podataka[1] (eng. *data warehouse*), jezero podataka (eng. data lake).

Baza podataka je strukturno organizovana kolekcija informacija, organizovanih tako da im se može brzo pristupati, te da se mogu brzo i lako pretraživati i ažurirati. Skladište podataka predstavlja vid centralizovanog spremišta podataka jedne kompanije, koji su prikupljeni iz više različitih izvora. Suština ovog tipa skladišta je olakšavanje procesa poslovne inteligencije (eng. *Business intelligence - BI*), tj. izvršavanje upita i vršenje analitičkih procesa nad velikim količinama podataka, koji su pre smeštanja u skladište struktuirani i prečišćeni. Sa druge strane, jezero podataka predstavlja centralizovano spremište takođe velikih količina podataka s tim što se za razliku od skladišta podataka (*data warehouse-a*) u jezero smeštaju i struktuirani i nestruktuirani podaci.



1. Iako se pojam skladište podataka koristi uglavnom u kontekstu spremišta podataka, bilo kog tipa, ovde se misli konkretno na jedan tip spremišta, čiji je (subjektivno) najprirodniji prevod na srpski jezik upravo skladište podataka.

PRAKTIČNI
AI PROJEKTI



RAZVOJ AI PROIZVODA

Projektima iz oblasti veštačke inteligencije uglavnom prethodi jedna istraživačka faza. Iako je sama oblast izuzetno interesantna, potrebno je biti potpuno upućen u mogućnosti i ograničenja alata i algoritama veštačke inteligencije. Oblast se konstantno razvija i usavršava, tako da je nivo progresa izvan svih očekivanja. Zato i jeste važno pre započinjanja projekta, osvežiti trenutna znanja, posavetovati se sa kolegama iz struke kao i konsultovati nedavno objavljene naučne radove.

Važno je napomenuti da je razvoj AI proizvoda zahtevna investicija, kako u pogledu vremena i ljudi, tako i neophodne infrastrukture i računarskih resursa. Sa druge strane, razvoj takvih projekata je izuzetno interesantan, uložena sredstva mogu se vratiti višestruko i što je možda najbitnije, otvaraju se vrata da će vašem proizvodu dati veliku prednost u odnosu na konkurenciju.

Nakon inicijalne ideje, pristupamo narednim fazama razvoja AI proizvoda:

- **Definisanje projekta** (engl. *project scoping*)
- **Upravljanje podacima** (engl. *data management*)
- **Razvoj ML modela** (engl. *ML model development*)
- **Spuštanje modela u produkciju** (engl. *ML model deployment*)
- **Nadzor performansi rada modela i održavanje** (engl. *monitoring and maintenance*)

Ono što AI proizvode razlikuje od klasičnih softverskih proizvoda jeste činjenica da ove faze nisu spregnute samo u jednom smeru. Rezultat svakog koraka može se odraziti na vraćanje na neki od prethodnih. Na primer, ukoliko se primeti da u radu modela u produkciji metrike ne zadovoljavaju vrednosti definisane prilikom planiranja projekta, u zavisnosti šta je uzrok tome, projekat se može vratiti u fazu pripreme podataka ili razvoja novog modela.

DEFINISANJE PROJEKTA

Sa ovom fazom nastavljamo nakon inicijalne ideje. Razmatra se izvodljivost projekta u okvirima trenutnih tehnoloških ograničenja, potrebnog kadra, raspoloživošću podataka. Takođe je izrazito bitno postaviti kriterijume prema kojima ćemo meriti uspešnost projekta, određivanje KPI-eva (engl. *Key Performance Indicators*). Definisanje dobrih metrika pomaže pri praćenju napretka, kao i za blagovremeno upozoravanje na neke potencijalne barijere u daljem razvoju.

UPRAVLJANJE PODACIMA

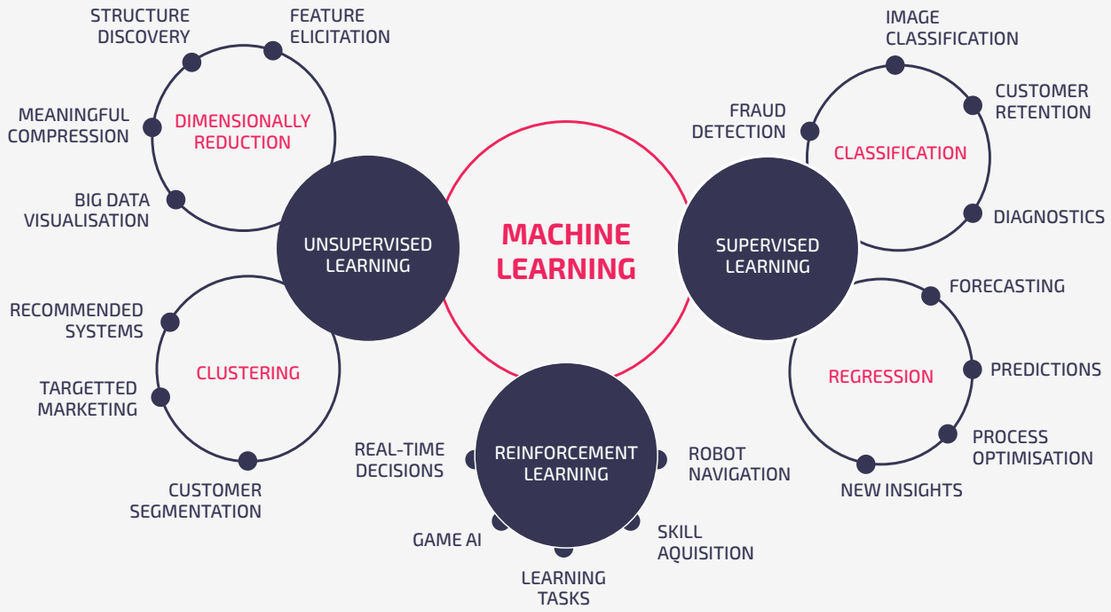
U ovoj fazi je izuzetno važno adresirati dva procesa: prikupljanje podataka i njihova organizacija. Za razvoj uspešnog AI proizvoda, kvalitetni (relevantni) podaci su ključni. Ovo je možda i jedan od najčešćih uzroka neuspešnosti nekog AI projekta (neodgovarajući podaci ili potpuno odsustvo istih).

U zavisnosti od toga koji problem rešavamo, postoji mogućnost da postoje podaci koji su možda već dostupni negde na internetu. U tom slučaju treba dobro razmisliti i proceniti da li ti podaci predstavljaju reprezentativan skup podataka problema koji želimo da rešimo. Da li su podaci sveži, odnosno da li je vremenska komponenta ključna za opisivanje našeg problema, u suprotnom možemo koristiti “zastarele” podatke.

Ako na kraju zaključimo da ne postoje javno dostupni podaci koji nam mogu biti od koristi, potrebno je što ranije formirati procese za prikupljanje istih. Zlatno pravilo je da je bolje prikupljati više podataka nego što je potrebno, nego obrnuto. Praksa pokazuje da retko šta od prikupljenih podataka, ostane neiskorišćeno na kraju, makar to bilo i mesecima, godinama kasnije.

RAZVOJ ML MODELA

U zavisnosti od prirode problema postoje različite grupe algoritama veštačke inteligencije koji su nam na raspolaganju.



Slika 3.1. - Izvor: https://pbs.twimg.com/media/ETNlvGZUYAI2iF5?format=jpg&name=large

Nakon izbora jednog, a često i više algoritama, pristupamo procesu njihovog treniranja na podacima sa kojima raspolažemo. U prvoj fazi podatke delimo u dve grupe, podatke za trening i podatke za validaciju. Podatke za trening koristimo da bismo trenirali i optimizovali pojedinačne algoritme. Svaki od algoritama ima neke (hiper)parametre kojima se kontroliše i upravlja rad algoritma. Nakon podešavanja parametara za svaki od algoritama pojedinačno, pristupa se njihovom međusobnom poređenju korišćenjem validacionog skupa. U zavisnosti od toga koju smo metriku izabrali da bude dominantna, što opet zavisi od specifičnog slucaja, vršimo odabir modela sa najboljim performansama.

SPUŠTANJE MODELA U PRODUKCIJU

Pod “spuštanjem modela u produkciju”, podrazumevamo proces prilikom kog naš model (koji pojednostavljenja radi, možemo da zamislimo kao jedan ili kao grupu fajlova) stavljamo na produkcioni server gde je dostupan glavnoj aplikaciji. Pored modela u produkciji, potrebno je podići i zasebne servise koji će motriti na rad modela, beležiti novopristigle podatke, itd.

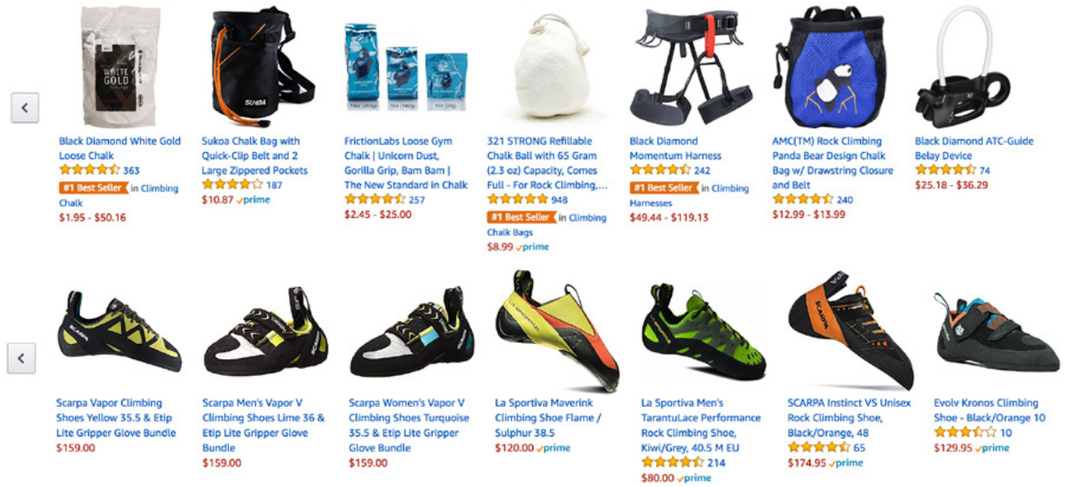
NADZOR PERFORMANSI RADA MODELA I ODRŽAVANJE

Kad je model konačno u produkcionom okruženju, tad tek započinje “pravi” životni ciklus jednog AI proizvoda. Nakon ove tačke započinje proces prikupljanja realnih podataka i validacija sa trenutno aktuelnim modelom. Prilikom tog procesa beležimo performanse rada modela (broj zahteva po sekundi, tačnost predikcije itd.). Nakon određenog vremena pristupa se re-treniranju modela, gde se novoprikupljeni podaci koriste za ponovni trening modela. Neretko se prave kombinacije podataka za trening, sa različitim udelom starih i novih podataka i naknadno odlučuje koji će od tih modela završiti nazad u produkciji.

PRIMERI GOTOVIH AI PROIZVODA

SISTEMI PREPORUKE

Sistemi preporuke (engl.) danas predstavljaju jedan od nezaobilaznih AI proizvoda bilo da se koriste za preporuku proizvoda, usluga, muzike, filmova, vesti, itd. Praktično je nemoguće naći danas veliki sistem koji pruža neki vid usluge širem krugu korisnika, a da ne uključuje sisteme preporuke.



Izvor: Amazon website // Preporuke na proizvod iz sekcije Free climbing shoes

Problem: Korisnik je došao da koristi vaše usluge, kao što smo već naveli to može da bude slušanje muzike, gledanje filmova, izbor sportske opreme, itd. Usko grlo vašeg servisa je uglavnom limitirano vreme koje je korisnik spreman da posveti vašoj aplikaciji/portalu itd. Za pronalaženje odgovarajućeg sadržaja, artikala ima previše, a vreme je ograničeno na nekoliko sekundi ili minuta. Ako korisnik u zadatom vremenskom okviru ne pronađe šta mu je potrebno, napušta vaš servis i prelazi na neki drugi.

Rešenje: Izrada sistema preporuke drastično skraćuje vreme pretrage do željenog proizvoda. Koriste se informacije o ponašanju svih korisnika na sajtu, njihova istorija pregleda, kupovine pa i vreme gledanja pojedinih artikala. Sve se radi sa ciljem da se unapred uoče obrasci ponašanja, da se izvrši klasterizacija i vrše preporučivanja proizvoda ili usluga koje su slični onome što trenutno gledate ili predikcija nečega što bi vam možda bilo interesantno da istražite. Pošto su sistemi preporuke nastajali u relativno ranijim fazama razvijanja vestačke inteligencije, tako imamo i dosta različitih implementacija koje koriste neke od osnovnijih algoritama sve do najnaprednijih koji koriste duboke neuronske mreže.

Izvor: NASA https://www.dpreview.com/news/1345517241/nasa-shares-high-resolution-satellite-images-of-california-s-camp-wildfire

DETEKCIJA ŠUMSKIH POŽARA

Usled globalnih klimatskih promena, šumski požari su sve češći i njihove posledice su sve drastičnije. Ranije su bili izolovani slučajevi, dok se danas tretiraju kao standardna pojava. Jedan od savremenih pristupa ranoj detekciji požara koristi AI sisteme nadzora satelitskih snimaka oblasti od interesa.

Problem: Prilikom letnjih meseci, aktivno se prate oblasti od interesa korišćenjem satelitskih snimaka. Proces upoređivanja snimaka je izuzetno zahtevan da bi se radio manuelnim putem. Tradicionalne metode upoređivanja slika "standardnim" softverskim alatima su davale mnogo lažnih uzbuna, ili su propuštale da prijave velike požare usled gustog dima koji se nadvijao nad nekom regijom.

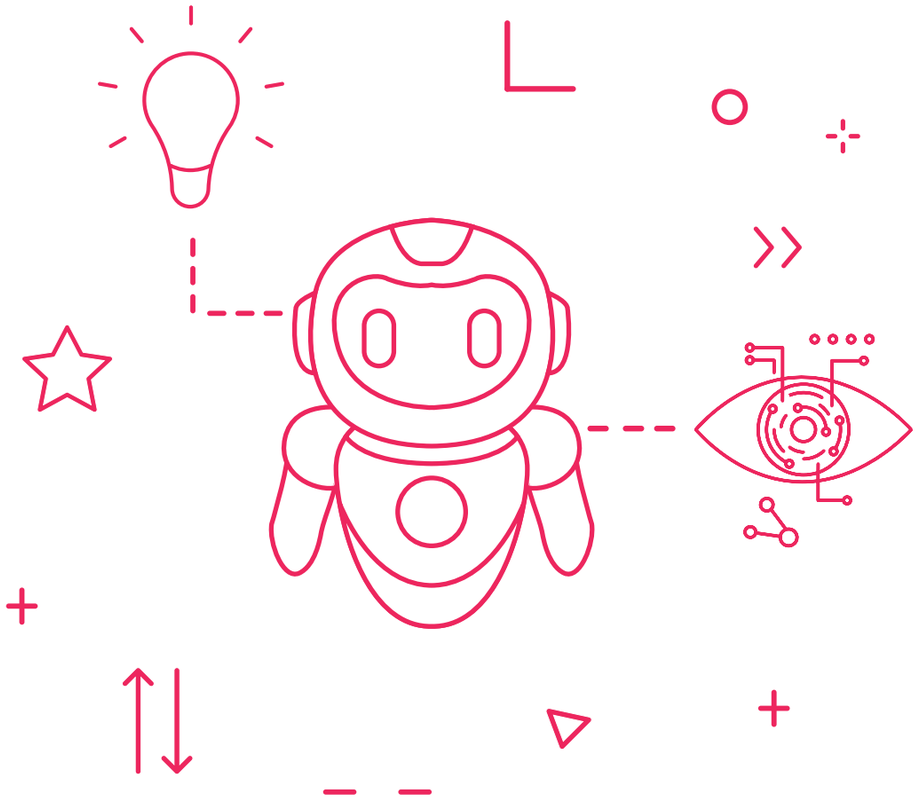
Rešenje: Upotrebom najsavremenijih CNN arhitektura (engl. *convolutional neural networks*), dobili smo mogućnost da treniramo modele na velikoj količini podataka koji su do sada snimljeni. Dodatno, možemo da zahvalimo ranijoj tehnologiji beleženja satelitskih snimaka koja je već mnogo godina u upotrebi. Sa sve više podataka, modeli za detekciju postaju sve precizniji sa sve manje lažnih uzbuna i što je svakako još bitnije, manje propusta kad su veći požari u pitanju.

CHATBOT

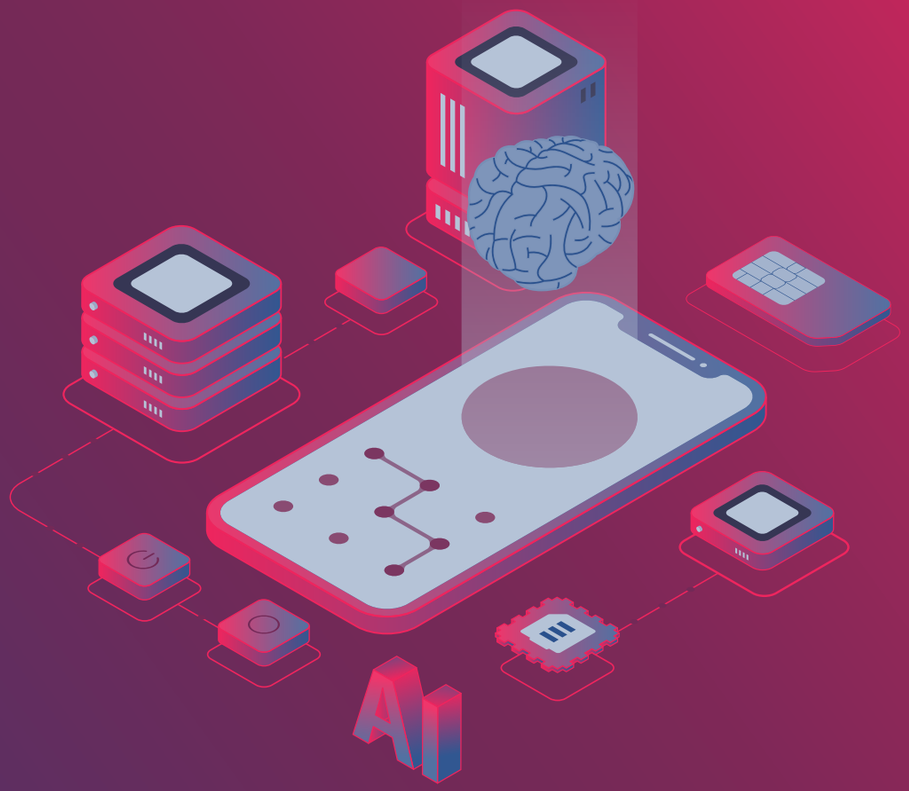
Istraživanja su pokazala da postojanje chatbot-a na sajtu/aplikaciji drastično unapređuje odnose sa finalnim korisnicima. Sa druge strane od izuzetne su pomoći da kanališu brojne zahteve klijenata ka ograničenim timovima kompanija koji su zaduženi za podršku korisnincima.

Problem: Korisnicima aplikacije ili vebsajta su potrebne jasne i pravovremene informacije. Najbolji scenario je kad bismo mogli odmah da stupimo u kontakt sa osobom iz kompanije koja ima potrebna znanja da adresira naš problem. Međutim, u zavisnosti od problema, potrebna je različita ekspertiza za njegovo rešavanje. Sa druge strane, postoji dosta korisničkih obraćanja koja su u suštini veoma slična. Tada je efikasnije preusmeriti korisnika na odredjenu sekciju sajta odnosno aplikacije gde može pronaći više informacija.

Rešenje: Chatbot je jedan kompleksan sistem čiji razvoj uključuje multidisciplinarne timove iz oblast lingvistike, veštačke inteligencije itd. Posebna grana AI-a, odnosno razumevanje prirodnih jezika (engl. *Natural Language Processing*) je odgovorna za razvoj ovih sistema. Ovakav sistem može da automatizuje značajan deo procesa u sprovodjenju korisnika do neophodnih odgovora na njihova pitanja. Klijent ima utisak da vrši konverzaciju sa osobom iz korisničke podrške. U početnim fazama, na osnovu postavljenih pitanja i odgovora na istih, vrši se upućivanje klijenta na sekcije sajta gde je moguće naći više informacija. U kasnijim fazama, ako su problemi konkretniji i specifičniji, vrši se preusmeravanje na osobu iz tima korisničke podrške sa najpotrebnijim znanjima za rešavanje problema. Ovim procesom se rasterećuju timovi od repetitivnih odgovora na često postavljana pitanja.



UTICAJ AI
NA DRUŠTVO



Kako je istaknuto u prethodnom tekstu, AI ima veliki potencijal i primenu u svakodnevnom životu ljudi. Samim tim, ona ima i veliki uticaj na društvo i najvažniji aspekt iste jeste humani - etika, moral, transparentnost i razumevanje modela i procesa. Ukratko, primena AI mora da služi za moralno dobro društva da bi se nazvala etičnom. Ovaj humani aspekt AI je predmet istraživanja, ali i svakodnevnica inženjera i članova timova koji razvijaju modele sa ciljem "spuštanja" istih u produkciju. Mi ćemo se, za potrebe ovog dokumenta, fokusirati na tri aspekta uticaja AI na društvo - zaštita podataka, etička upotreba istih i zaštita životne sredine.

ZAŠTITA PODATAKA

Kako Evropska komisija u vodiču o etici i zaštiti podataka (European Commission, "Ethics and Data Protection", 2021) navodi: "Zaštita podataka je i centralno pitanje istraživačke etike u Evropi, i osnovno ljudsko pravo".

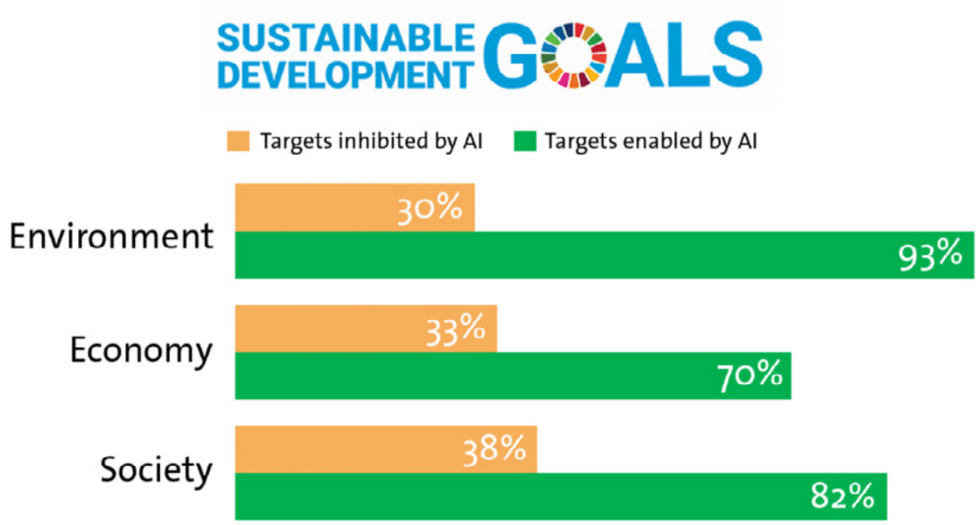
S obzirom na to da se podaci koje ostavljamo na internetu nazivaju našim digitalnim otiskom prsta, jasno je zašto je ovaj citat važan i istinit. Rigorozni zakoni o zaštiti podataka su potrebni i u istraživačkim i poslovnim scenarijima. Bez obzira na scenario, individue imaju pravo da znaju da se njihovi podaci koriste i kako se oni koriste. Posebnu pažnju treba obratiti na istraživanja i poslovne scenarije koja uključuju posebne kategorije podataka (ranije poznate kao osetljivi podaci, odnosno osnovni lični podaci, genetički podaci, itd). U te primene spada i AI, jer operacije obrade padataka tu mogu predstavljati veći rizik za prava i slobode subjekata podataka. Sve veći uticaj ovih i drugih novih tehnologija u našem svakodnevnom životu i aktivnostima ogleda se u Opštoj uredbi za zaštitu podataka (GDPR, "*General Data Protection Regulation*", 2016;). Prema tome, zaključujemo da bilo koja primena AI mora biti u skladu sa ovom uredbom. Međutim, činjenica da je primena zakonski dozvoljena ne znači nužno da će se smatrati etičnom.

ETIČKA UPOTREBA PODATAKA

Kako možemo biti sigurni da etički koristimo posebne kategorije podataka prilikom primene AI? Ovo nije trivijalno pitanje, međutim, s obzirom na to da je etika definisana kao set pravila definisanih od strane spoljašnjeg izvora, možemo se voditi vodiljama koje navodi Evropska komisija u vodiču o etici i zaštiti podataka. EU je (*European Commission*, "Ethics guidelines for trustworthy AI", 2019;) objavila skup smernica za izgradnju verodostojne veštačke inteligencije koja uključuje etičke stubove kao što su pravičnost, transparentnost i robusnost, pored odgovornosti, privatnosti, usaglašenosti i zakonitosti. U skladu sa principom pravičnosti i transparentnosti, metode prikupljanja podataka se pregledaju i proveravaju za adekvatnost podataka (npr. podaci koji nedostaju za određene grupe) i pristrasnost da bi se pripremili za primenu specifičnih tehnika ublažavanja u fazi pripreme podataka. U skladu sa principom privatnosti i upravljanja podacima, izvori podataka za lične informacije (engl. Personally Identifiable Information, PII) se moraju zabeležiti i opravdati. Nakon ovog koraka koji nazivamo mapiranjem posebnih izvora podataka, mora se odlučiti kako ćemo se sa istima nositi. Dve opcije koje se najčešće koriste jesu pseudonimizacija podataka (maskiranje PII i de-identifikacija) i upotreba sintetičkih podataka. Ostali principi se odnose na širu upotrebu AI i nisu predmet ovog teksta (vidi dodatne resurse, 2. stavka).

ZAŠTITA ŽIVOTNE SREDINE

Kao što je spomenuto, AI mora da služi za moralno dobro društva da bi njena upotreba bila etična. Dominantni primer takvih suštinskih moralnih dobara su ciljevi održivog razvoja UN (SDG, Sustainable Development Goals) (United Nations, "*The Sustainable Development Goals*", 2019). Iako AI ima negativan uticaj na neke, kao deo ovog skupa ciljeva, oni sa temom zaštite životne sredine su oni ciljevi gde AI ima najpozitivniji uticaj (Slika 4.1), ali su i zanimljivi sa etičke strane veštačke inteligencije jer se mogu shvatiti kao nešto najbliže konsenzusu čovečanstva u smislu moralnih ciljeva.



Slika 4.1: Uticaj AI na postizanje svakog SDG. Izvor: Ricardo Vinuesa, "Opportunities and challenges of AI to achieve the Sustainable Development Goals", 2020.

Primeri upotrebe veštačke inteligencije za zaštitu životne sredine mogu se podeliti na (*Capgemini Research Institute: Climate AI How artificial intelligence can power your climate action strategy,* 2020) proizvodnju i distribuciju energije, proizvodne operacije i upotrebu proizvoda.

PROIZVODNJA I DISTRIBUCIJA ENERGIJE

Primer upotrebe AI u ovom sektoru jeste praćenje emisija gasova staklene bašte na industrijskim lokacijama. ExxonMobil se udružio sa MIT-jem da dizajnira i razvija dubokomorske AI robote kako bi povećali mogućnost detekcije curenja nafte. Ova primena AI ima veliki uticaj obzirom da je curenje nafte odgovorno za skoro polovinu nafte koja se ispušta u okean svake godine (SparkCognition, "*Top 4 Real-World AI Applications In The Oil And Gas Industry*", 2020). Još jedan primer jeste upotreba AI za poboljšanje energetske efikasnosti. Kako Google piše, potencijal upotrebe AI u svrhe efikasnijeg korišćenja obnovljivih izvora energije je veliki. Google je postigao svoj cilj prestanka emitovanja ugljen-dioksida 2018. godine (*Google, "Meeting our match: Buying 100 percent renewable energy*", 2018;) i sada, u saradnji sa Deepmind-om, koriste AI za predviđanje obrazaca vetra za poboljšanje prinosa vetrenjača (Deepmind, "*Machine learning can boost the value of wind energy,*" 2019;). Takođe, Google koristi AI modele da smanji energetsko opterećenje svojih centara podataka, smanjujući troškove energije za hlađenje za 40% (Wired, "Google's DeepMind trains AI to cut its energy bills by 40%", 2017;). Konačno, korišćenje međusobno povezanih tehnologija kao što su električna autonomna vozila i pametni uređaji sa veštačkom inteligencijom koja reaguje na potražnju može pomoći u poboljšanju emisije gasova u gradovima i smanjenju zavisnosti od fosilnih goriva.

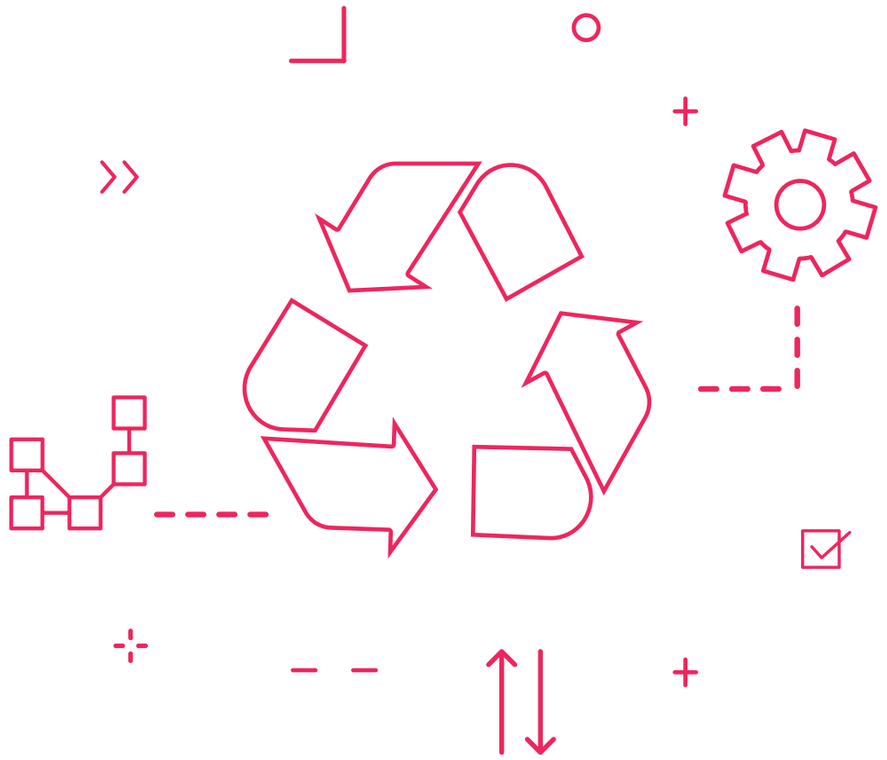
PROIZVODNE OPERACIJE

Kao što smo već spomenuli, upotreba AI modela za upravljanje rutama može pomoći u smanjenju potrošnje goriva i emisije ugljenika. Jedan primer jeste Locus.sh, kompanija koja koristi AI baš u ove svrhe (Locus.sh, "*How the World's Best Route Optimization Engine Works*," 2018;). AI može pomoći u predviđanju i izbegavanju mehaničkih kvarova, kao u slučaju General Motors koji koriste modele za klasifikaciju slika za otkrivanje kvarova komponenti. Ovo je važan primer upotrebe AI za prediktivno održavanje i otkrivanje kvarova u proizvodnim operacijama (iFlexion, "*Industries to Be Transformed by Machine Learning for Image Classification*", 2018;). AI takođe povećava poljoprivrednu efikasnost. Koristeći prediktivnu analitiku, poljoprivrednici mogu da koriste ispravnu količinu đubriva i vodnih resursa da bi proizveli najveći prinos. Rizik se takođe može svesti na minimum analizom vremenskih podataka, štetočina i ekstremnih vremenskih događaja. Poljoprivreda je veliki izvor emisije ugljenika, što čini veštačku inteligenciju moćnim oružjem u borbi protiv ugljenika (Datamation, "*How the Agriculture Industry is Using AI*", 2021;).

UPOTREBA PROIZVODA, PONOVNA UPOTREBA I RECIKLAŽA

AI ima značajan potencijal upotrebe za efikasno upravljanje otpadom, segregaciju i reciklažu kako bi se smanjila potreba za resursima. Na primer, ZenRobotics koristi AI za procenjivanje količine otpada i za efikasno razdvajanje otpada da bi poboljšali proces reciklaže (*Recycling Product News*, *"ZenRobotics' AI-based robotic waste sorting technologies help Remeo to build next-generation MRF"*, 2020;).

Ove primene AI ilustruju da ona može pokrenuti tranziciju ka ekonomiji sa nultom emisijom ugljenika, međutim, ako se vratimo na sliku 4.1, primećujemo da AI ima i negativan uticaj na SDG. Usvajanje AI rešenja mora da bude praćeno transparentnošću o njegovoj negativnoj strani – ugljenični otisak kompleksnih AI rešenja je značajan. Uvid u otisak AI rešenja i drugih IT alata je neophodan da bi se isti koristili u svrhu zaštite životne sredine.



DODATNI RESURSI

1. S obzirom na to da je oblast veštačke inteligencije izuzetno dinamična potrebno je pronaći dobar izvor "svežih", a relevantnih informacija. Nešto što ispunjava te uslove je nedeljni blog The Batch, koji DeepLearning. AI objavljuje svakog četvrtka.
   https://read.deeplearning.ai/the-batch/
2. Etička primena AI sa posebnim osvrtom na alate otvorenog tipa jeste predmet istraživanja AI tima u Sogeti NL. Za više informacija o tome kako se možete postarati da je vaša izrada i poslovna primena AI pravična i kvalitetna, pročitajte sledeći članak.
   https://www.sogeti.com/ai-for-qe/section-9-trust-ai/chapter-1/
3. Primeri gde je AI uspešnija od ljudi.
   https://www.businessinsider.com/robots-surpassed-humans-at-these-tasks-in-past-decade-2019-11
4. Ograničenja AI.
   https://medium.com/codex/how-far-can-artificial-intelligence-go-the-8-limits-of-machine-learning-383dd9b2f7bd
5. Praktikum iz mašinskog učenja, pisan za studente FTN, UNS.
   https://www.project-benefit.eu/eplatform/?courses=33&download=MU_praktikum.pdf
6. Knjiga "Veštačka inteligencija", Predraga Janičića i Mladena Nikolića, koja se između ostalog koristi na predavanjima iz kursa "Veštačka inteligencija" na MATF-u. Dostupna je i u besplatnom elektronskom izdanju.
   http://poincare.matf.bg.ac.rs/~janicic//books/VI_B5.pdf
7. Besplatan kurs jednog od najvećih imena veštačke inteligenije u svetu, Andrew Ng (suosnivač i rukovodilac Google Brain odeljenja, takođe i jedan od osnivača Coursera-e).
   https://www.coursera.org/learn/ai-for-everyone
8. "Elements of AI" je serija besplatnih online kurseva kreirana od strane Univerziteta u Helsinkiju.
   https://www.elementsofai.com/
9. O tipovima procesora (CPU, GPU, TPU) i njihovo poređenje, Google AI & Machine Learning Blog.
   https://cloud.google.com/blog/products/ai-machine-learning/what-makes-tpus-fine-tuned-for-deep-learning

NASLOV: UVOD U VEŠTAČKU INTELIGENCIJU
GODINA: 2022
AUTORI: MLADEN JOVANOVIĆ, MLADEN CANOVIĆ, DANICA DESPOTOVIĆ, TIJANA NIKOLIĆ

DIZAJN: KSENIJA SAVIĆEVIĆ
IZDAVAČ: INICIJATIVA "DIGITALNA SRBIJA" (DSI)
ZA IZDAVAČA: NEBOJŠA BJELOTOMIĆ, DIREKTOR PROJEKTA PREDUZMI IDEJU

www.preduzmi.rs

Naslov: Uvod u veštačku inteligenciju

Autori: Mladen Jovanović, Mladen Canović, Danica Despotović, Tijana Nikolić

Izdavač: Inicijativa „Digitalna Srbija“

Štampa: Inicijativa „Digitalna Srbija“, bulevar Milutina Milankovića 11a, Beograd

Tiraž: 10

Mesto izdavanja: Beograd

Godina izdavanja: 2024.

ISBN 978-86-82900-07-8

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

004.8(0.034.2)

UVOD u veštačku inteligenciju [Електронски извор] / [autori Mladen Jovanović ... [et al.]]. - Beograd : Inicijativa "Digitalna Srbija", 2024 (Beograd : Inicijativa "Digitalna Srbija"). - 1 USB fleš memorija ; 1 x 2 x 7 cm

Sistemski zahtevi: Nisu navedeni. - Tiraž 10. - Sadrži bibliografiju.

ISBN 978-86-82900-07-8

1. Јовановић, Младен, 1974- [аутор]
a) Вештачка интелигенција

COBISS.SR-ID 140926985