

---

*JEZIČNI AGENTI: RAZVOJ KOMUNIKACIJSKIH  
PARTNERA POMOĆU VELIKIH JEZIČNIH  
MODELAA*

Benedikt Perak

## SADRŽAJ

---

1	Uvod: Komunikacija i razvoj civilizacije .....	6
2	Povijest i evolucija komunikacijskih tehnologija .....	9
2.1	Komunikacija kao osnova razvoja civilizacije .....	9
2.2	Usmene predaja i živi arhivi .....	9
2.2.1	Važnost usmene tradicije.....	9
2.2.2	Kolektivna imaginacija i identitet .....	10
2.2.3	Društvene funkcije usmene tradicije.....	11
2.3	Pismenost kao prozor u složene sisteme znanja .....	11
2.3.1	Različite staze do pisma.....	11
2.3.2	Društvena elita i sporost širenja .....	12
2.4	Tiskarski stroj: Prekretnica u širenju ideja .....	12
2.4.1	Masovna proizvodnja i demokratizacija” znanja .....	13
2.4.2	Reformacija i znanstvena revolucija .....	13
2.4.3	Globalno širenje i kulturni utjecaj .....	13
2.4.4	Preteča informacijskog društva .....	14
2.4.5	Nasljeđe za daljnji razvoj komunikacije.....	14
2.5	Telegraf i telefon: Početak elektroničke komunikacije .....	14
2.5.1	Telegraf: nove brzine i novi horizonti .....	14
2.5.2	Telefon: intimna revolucija na žici” .....	15
2.5.3	Preoblikovanje društvene dinamike .....	15
2.5.4	Impuls za buduće komunikacijske revolucije .....	16
2.6	Radio i televizija kao katalizatori masovnog iskustva .....	16
2.6.1	Uspon radija i glas” zajedničke stvarnosti .....	16
2.6.2	Televizija i vizualna dimenzija zajedničke imaginacije .....	16
2.6.3	Masovna komunikacija i homogenizacija kulture.....	17
2.6.4	Politička sfera i elektroničko javno mnjenje” .....	17
2.6.5	Komunikacija kao temelj civilizacije: pogled unaprijed .....	17
2.7	Internet i digitalna revolucija .....	19
2.7.1	Proširenje komunikacijskih kanala .....	20
2.7.2	Suradnja i kolaborativni alati .....	20
2.7.3	Ekspanzija društvenih mreža.....	20
2.7.4	Digitalna Plemena i Trgovi Znanja: Anatomija Uspješnih Online Zajednica.....	21
2.7.5	Porozne Membrane: Kad Digitalno Postane Stvarno (i Obratno) .....	23
2.7.6	Digitalna revolucija i mreža globalne komunikacije .....	24

2.7.7	Digitalna Dvojnost: Obećanje Demokratizacije i Sjena Novih Nejednakosti .....	25
2.7.8	Sjene u digitalnom edenu: privatnost, manipulacija i cijena povezanosti.....	26
2.8	Začeci Autonomne Komunikacije: Algoritamski Šapati i Prvi Digitalni Sugovornici ....	28
3	Svijet Velikih Jezičnih Modela: Od Teorije do Tehnologije .....	30
3.1	Uvod: Ulazak u Doba Velikih Jezičnih Modela .....	30
3.2	Anatomija LLM-a: Ključne Tehnologije i Arhitekture .....	33
3.3	Životni Ciklus Modela: Od Podataka do Primjene .....	36
3.3.1	Pre-treniranje (Pre-Training): Stvaranje Temeljnog Znanja.....	36
3.3.2	Fino Podešavanje i Poravnanje (Fine-Tuning & Alignment): Prilagodba Svrsi .....	39
3.3.3	Optimizacija za Stvarni Svijet: Efikasnost (Efficiency) .....	41
3.3.4	Mjerenje Uspjeha: Evaluacija (Evaluation).....	44
3.3.5	Model u Akciji: Inferencija (Inference) .....	48
3.4	Izazovi i Ograničenja: Sjene u Digitalnom Ogledalu .....	52
3.4.1	Sigurnost, Pouzdanost i Zloupotraža .....	52
3.4.2	Pristrandost, Privatnost i Etika Podataka: Tamna Strana Digitalnih Ogledala .....	54
3.4.3	Resursna Intenzivnost: Cijena Inteligencije.....	57
3.5	Transformacija Komunikacije: Primjene i Perspektive .....	59
3.6	LLM: Sljedeća Stanica Evolucije? .....	63
4	DEKONSTRUKCIJA JEZIKA U DOBA AI: OD SIMBOLA DO DRUŠTVENE STVARNOSTI .....	65
4.1	Jezik kao Struktura Značenja: Simboli, Odnosi i Mentalne Mape .....	65
4.1.1	Strukturalistički Odjaci u Arhitekturi LLM-ova: Značenje kao Relacija.....	65
4.1.2	Semantički Trokut i Kognitivna Dimenzija: Izazov Značenja za AI .....	67
4.2	Jezik kao Scenarij: AI Agenti i Simulacija Društvenih Interakcija .....	70
4.3	Jezik kao Arhitektura Društva: Gradnja i Održavanje Socijalnih Struktura kroz Komunikaciju .....	73
4.4	LLM-ovi kao Komunikacijski Akteri: Oblikovanje Društvenog Konteksta kroz Algoritamsku Interakciju .....	75
4.4.1	Jezik kao Ogledalo i Kalup Identiteta: Utjecaj AI Agenata na Naše "Ja" .....	77
4.4.2	Jezik, AI Agenti i (Re)konstrukcija Stvarnosti: Algoritamske Leće i Okviri Značenja	81
4.4.3	Jezik kao most ka apstrakciji: Stvaranje i dijeljenje nevidljivih svjetova .....	84
4.4.4	Jezik i Kolektivno Stvaranje: Od Društvenih Ugovora do AI Suradnje .....	86
4.4.5	AI Agenti kao Sugovornici: Više od Pukog Alata.....	90
5	POGON INTELIGENCIJE: PROCESORSKA SNAGA, AI HORIZONTI I RAĐANJE DIGITALNIH KOLEKTIVA .....	93
5.1	Motor Evolucije: Od Mooreovog Zakona do AI Revolucije.....	93
5.2	Širenje Horizonata: Od Uskih Zadataka do Digitalnih Kolektiva .....	95

5.3	Potraga za Dubljim Razumijevanjem: Emocionalna Inteligencija i AGI .....	98
5.4	Temeljni Zahtjevi: Ekosustav za Naprednu AI .....	99
5.5	Zaključak: Pogon za Budućnost Komunikacije .....	100
<b>6</b>	<b>OD JEZIČNOG MODELA DO KOMUNIKACIJSKOG PARTNERA: ARHITEKTURA I PRIMJENA AI AGENATA.....</b>	<b>102</b>
6.1	Programabilni Jezik: Uloga API-ja u Interakciji s LLM-ovima .....	102
6.2	Transformacija: Od LLM Jezgre do Funkcionalnog Agenta.....	103
6.3	Anatomija Modernog Komunikacijskog Partnera: Ključne Tehnike .....	104
6.3.1	Umjetnost Upravljanja: Prompt Inženjering – Kormilarenje Jezičnim Oceanom ..	104
6.3.2	Sistemska uputa.....	108
6.3.3	Utemeljenje u Stvarnosti: Retrieval-Augmented Generation (RAG).....	108
6.4	Korištenje Alata (Tool Use): Od Riječi do Djela.....	112
<b>7</b>	<b>IZGRADNJA KOMUNIKACIJSKIH PARTNERA: PRIMJERI I STUDIJE SLUČAJA .....</b>	<b>113</b>
7.1	Digitalni Sugovornici: Chatbotovi i Virtualni Asistenti u Doba LLM-ova .....	113
7.2	Primjene Koje Mijenjaju Igru: Gdje Agenti Stvaraju Vrijednost?.....	114
7.2.1	Revolucija u Korisničkoj Podršci i Marketingu.....	114
7.2.2	Transformacija Obrazovanja i Personalizacije Sadržaja .....	116
7.2.3	Proširenje Primjenjivosti: LLM Agenti u Zdravstvu, Poslovnim Procesima i Kreativnim Industrijama .....	117
7.2.4	Paradigma Personalizacije Sadržaja: Transverzalna Primjena AI Algoritama .....	118
7.2.5	Izvan Funkcionalnosti: Početak Širih Razmatranja.....	119
<b>8</b>	<b>8 DIGITALNI SUPUTNICI: VIZIJA SVAKODNEVICE U DOBA SVEPRISUTNIH AGENATA (cca. 2030-2035) .....</b>	<b>120</b>
8.1	Uvod: Od Današnjice do Sutrašnjice – Projekcija Trendova.....	120
8.2	Dan u Životu (cca. 2033): Narativna Skica .....	121
8.2.1	Jutarnja Simfonija: Personalizirano Budenje i Koordinacija .....	121
8.2.2	Radni Tokovi i Učenje Budućnosti: Agenti kao Suradnici i Mentorji .....	123
8.2.3	Društvena Mreža i Povezanost: Agenti kao Posrednici .....	125
8.2.4	Automatizirana Svakodnevica i Neočekivani Trenuci .....	126
8.3	Dešifriranje Vizije: Ključni Trendovi i Tehnološki Pokretači .....	128
8.3.1	Ambientna Inteligencija i Sveprisutnost Agenata .....	128
8.3.2	Proaktivnost, Autonomija i Delegiranje Odluka .....	128
8.3.3	Hiper-Personalizacija .....	129
8.3.4	Sinergija Rojeva .....	129
8.3.5	Čovjek-AI Simbioza: Ne samo Alat, već Partner .....	129
8.4	Glasovi Budućnosti: Perspektive Mislilaca i Istraživača .....	130

8.5	Društvo u Transformaciji: Naznake Novih Norma i Izazova.....	131
8.6	Zaključak Poglavlja: Između Utopije i Distopije – Otvorena Pitanja .....	133
9	Reference.....	133

# 1 UVOD: KOMUNIKACIJA I RAZVOJ CIVILIZACIJE

## Verzija

Od prvih artikuliranih glasanja s ognjišta ljudske svjesnosti do složenih simfonija riječi koje ispunjavaju naše digitalne prostore, komunikacija je bila sveprisutna sila koja pokreće kotače civilizacije. Poput tkiva koje povezuje pojedince u zajednici, ljepila koje drži kulturu na okupu i gorivo koje napaja motor inovacija. Jer, bez sposobnosti dijeljenja misli, iskustava i znanja, bili bismo tek bića svedena na instinkte. Tek usmena predaja, ta prva velika **komunikacijska tehnologija**, pretvorila je prolazna sjećanja u žive arhive, utkivala ih u pjesme, mitove i rituale, prenoseći mudrost i identitet kroz generacije poput genetskog koda kulture. Pojavom sustava pisma, riječi su otisnute na glini, papirusu i papiru, omogućivši akumulaciju znanja, razvoj apstraktne misli i izgradnju složenih administrativnih i filozofskih sustava koji su definirali carstva i epohe. Svaka daljnja nova komunikacijska tehnologija – od Gutenbergovog tiskarskog stroja koji je demokratizirao znanje, preko telegraфа i telefona koji su srušili prostorne barijere, do radija, televizije i interneta koji su stvorili globalno selo – predstavljala je novo poglavlje u ovoj neprestanoj evoluciji načina na koji razmjenjujemo značenje.

Danas stojimo na pragu, ili smo možda već zakoračili preko njega, u novo, zapanjujuće doba komunikacije obilježeno tehnološkom pojavom velikih jezičnih modela (eng. *Large Language Model* - LLM). Ovi napredni sustavi umjetne inteligencije, poput GPT (OpenAI, 2024b), Gemini (Google), Claude (Anthropic, 2024) ili Llama (Meta AI, 2024) istrenirani su na nezamislivim prostranstvima digitalnog teksta i koda – goleminim repozitorijima ljudskog jezika i znanja – prepoznavati nevjerojatno suptilne obrasce komunikacije. Tu se krije i njihova fascinantna dvojnost, koju je elegantno sažela lingvistica Milena Žic Fuchs (**Fuchs**), razlikujući znanje o svijetu (činjenice, koncepti) i znanje o jeziku (gramatika, stil, pragmatika). Moderni LLM-ovi besprijekorno barataju jezikom, generirajući tekstove koji su gramatički ispravni, stilski pogodni, činjenično točni i kontekstualno relevantni. Oni uvjerljivo oponašaju ljudsku jezičnu produkciju u specifičnim zadacima, ponekad dosežući razine performansi koje ih čine superiornima.

Dapače, kada tim sofisticiranim generatorima teksta definiramo ciljeve i ugradimo u šire informacijske sustave, s alatima za pristupom podacima i memoriji, oni postaju komunikacijski agenti – naši novi digitalni sugovornici, asistenti, suradnici, a možda čak i, kako sugerira naslov ove knjige, **komunikacijski partneri**. Već sada, oni preoblikuju krajolik ljudske interakcije: prevode jezike u stvarnom vremenu s neviđenom preciznošću, pokreću chatbotove koji pružaju personaliziranu korisničku podršku 24/7, pomažu znanstvenicima u analizi podataka, asistiraju programerima u pisanju koda, generiraju kreativne ideje i omogućuju nove oblike globalne suradnje i obrazovanja. Oni su poput univerzalnih prevoditelja i posrednika znanja, sa živom sposobnošću premošćivanja jezičnih, disciplinarnih i kulturnih barijera.

Ova knjiga je strukturirana kao putovanje koje započinje duboko u prošlosti, a završava pogledom u moguću budućnost interakcije čovjeka i stroja. Naš put započinje detaljnim pregledom povijesti komunikacijskih tehnologija (Poglavlje 2), od usmene predaje do digitalne revolucije. Razumijevanje kako su prethodne tehnologije oblikovale društvo i ljudsku spoznaju ključno je za procjenu potencijalnog utjecaja sadašnje AI revolucije.

Nakon postavljanja temelja, zaranjamo u srce tehnologije koja pokreće današnje komunikacijske agente: svijet velikih jezičnih modela (Poglavlje 3). Razotkrit ćemo njihovu anatomiju, prije svega arhitekturu Transformera i životnog ciklusa jezičnih modela – od prikupljanja golemih skupova podataka i pre-treniranja, preko finog podešavanja i ključnog

**Commented [BP1]:** Od prvih riječi izgovorenih u krugu zajednice do sofisticiranih dijaloga s umjetnom inteligencijom, komunikacija je bila i ostala temelj razvoja ljudske civilizacije. Kroz stoljeća, ona je omogućila dijeljenje znanja, izgradnju identiteta i oblikovanje zajedničkih vizija budućnosti. Usmena predaja stvarala je žive arhive povijesnog sjećanja, dok je pismenost otvorila vrata složenim sustavima znanja, osiguravajući temelje za širenje ideja i inovacija. Danas svjedočimo novom razdoblju komunikacije, obilježenom razvojem velikih jezičnih modela (LLM). Ovi napredni sustavi umjetne inteligencije, trenirani na goleminama podataka, sposobni su prepoznavati obrasce znanja o svijetu i znanja o jeziku (Žic Fuchs) i generirati prirodnji jezik na razinama koje oponašaju ljudsku interakciju. Kao komunikacijski agenti, LLM-ovi premošćuju jezične i kulturne barijere, omogućuju personaliziranu razmjenu informacija i stvaraju nove mogućnosti za globalnu suradnju.

Ova knjiga razmatra povijesnu evoluciju komunikacije, od usmenih tradicija i pisanih zapisova do digitalnog doba. Središnja tema je uloga velikih jezičnih modela u oblikovanju suvremenih komunikacija, njihov potencijal za unapređenje društvenih interakcija, kao i etički izazovi koje donose. Ključni ciljevi knjige uključuju: 1. Povezivanje tradicionalnih i suvremenih oblika komunikacije. 2. Analizu načina na koje LLM-ovi redefiniraju poimanje dijaloga, suradnje i znanja. 3. Refleksiju o društvenim i filozofskim implikacijama AI tehnologija.

Veliki jezični modeli (eng. *Large language models* - LLM) su napredni AI sustavi koji koriste duboke neuronske mreže za analizu, razumijevanje i generiranje prirodnog jezika. Njihova snaga leži u sposobnosti da prepozna obrasce i kontekst u velikim količinama podataka, što ih čini ključnim alatom za komunikacijske agente. Takvi sustavi pomažu u automatizaciji dijaloga, personalizaciji korisničkog iskustva i premošćivanju jezičnih prepreka, čineći ih neprocjenjivima u obrazovanju, poslovanju i društvenoj interakciji. LLM-ovi preuzimaju ulogu mosta između tradicionalne ljudske komunikacije i nove ere digitalnih dijaloga s primjenom u međukulturne suradnje, ali i postavlja pitanja o odgovornosti, pristranosti i etičkoj uporabi tehnologije.

Komunikacija ostaje ključni mehanizam koji povezuje ljudе i ideje, stvara institucionalne i kulturne obrasce. U ovom povijesnom trenutku, veliki jezični modeli donose priliku da taj mehanizam postane brži, inkluzivniji i moćniji nego ikad prije. Kroz ovu knjigu istražit ćemo kako se prošlost i budućnost susreću u doba umjetne inteligencije, oblikujući nove procese ljudske interakcije.

procesa poravnjanja (alignment) s ljudskim vrijednostima, sve do evaluacije performansi i izazova njihove implementacije u stvarnom svijetu. Nećemo zaobići ni tamniju stranu – inherentna ograničenja, rizike pristranosti, sigurnosne probleme i ogromnu resursnu intenzivnost ovih digitalnih divova.

S tehnološkim razumijevanjem kao osnovom, ulazimo u dublju, konceptualnu analizu odnosa jezika, društva i umjetne inteligencije (Poglavlje 4). Ovdje ćemo dekonstruirati jezik kao sustav značenja i društvene prakse, istražujući kako LLM-ovi, kao novi komunikacijski akteri, interveniraju u te procese. Promatrat ćemo kako simuliraju društvene interakcije, utječu na naš osjećaj identiteta i potencijalno (re)konstruiraju našu percepciju stvarnosti, djelujući kao moćne algoritamske leće. Posebnu pažnju posvetit ćemo ulozi jezika u kolektivnom stvaranju znanja i kulture, te kako AI agenti postaju novi sudionici u tim procesima – od suradnje u znanosti do ko-kreacije u umjetnosti.

Ne možemo govoriti o AI-u bez razumijevanja fizičkog pogona koji stojiiza nje (Poglavlje 5). Istražit ćemo neraskidivu vezu između eksponencijalnog rasta procesorske snage i napretka AI, povlačeći paralele s evolucijom ljudskog mozga. Razmotrit ćemo kako ova snaga omogućuje današnje LLM-ove i nagovještava horizonte budućih, još moćnijih sustava, uključujući potencijalne multi-agentske rojeve i dugoročnu potragu za općom umjetnom inteligencijom (AGI).

Nakon što smo postavili teorijske i tehnološke temelje, prelazimo na praktičnu izgradnju komunikacijskih partnera (Poglavlje 6). Ovo poglavlje detaljno opisuje kako se od temeljne LLM jezgre dolazi do funkcionalnog AI agenta. Istražit ćemo ključne tehnike poput prompt inženjeringu (umijeća upravljanja LLM-om), **Retrieval-Augmented Generation** (RAG) koja omogućuje agentima pristup i korištenje vanjskog, ažurnog znanja, te mehanizme za korištenje alata (Tool Use) koji agentima daju sposobnost djelovanja u digitalnom svijetu.

Kako ovi agenti izgledaju u praksi? Poglavlje 7 donosi konkretne primjere i studije slučaja, ilustrirajući kako se AI agenti koriste za transformaciju korisničke podrške, personalizaciju obrazovanja, unapređenje marketinških strategija, te kako pronalaze primjenu u zdravstvu, poslovnim procesima i kreativnim industrijama. Analizirat ćemo načine na koji stvaraju dodanu vrijednost, ali i nove izazove koje donose.

Konačno, u posljednjem poglavlju (Poglavlje 8), odvažit ćemo se na spekulativni, ali utemeljeni pogled u blisku budućnost (cca. 2030-2035). Kroz narativnu skicu dana u životu i analizu ključnih trendova poput ambientne inteligencije, hiper-personalizacije i sinergije agentskih rojeva, pokušat ćemo vizualizirati kako bi mogla izgledati svakodnevica prožeta sveprisutnim, proaktivnim i autonomnim digitalnim suputnicima. Razmotrit ćemo potencijal za istinsku čovjek-Al simbiozu, kao i društvene i etičke dileme koje takva budućnost nosi, balansirajući između utopijskih obećanja i distopijskih rizika.

Kroz cijelu knjigu provlače se ključni koncepti: evolucija komunikacije kao neprekinuti proces, LLM kao moćna, ali i ograničena tehnologija, AI agenti kao entiteti koji nadilaze puki alat i postaju komunikacijski partneri, te stalna potreba za kritičkim promišljanjem etičkih implikacija i odgovornim oblikovanjem budućnosti interakcije čovjeka i stroja. Naše putovanje ima za cilj pružiti čitatelju znanje o tome kako ovi sustavi funkcioniraju te dublje razumijevanje zašto su važni i kamo bi nas mogli odvesti.

Kroz cijelu povijest, komunikacija je bila ključni mehanizam kojim smo gradili mostove između umova, stvarali zajedničke stvarnosti i oblikovali institucije i kulture koje definiraju naš svijet.

Danas, na ovom povijesnom raskrižju, veliki jezični modeli i AI agenti koje oni pokreću nude nam priliku – ali i izazov – da taj mehanizam učinimo eksponencijalno bržim, globalno dostupnijim i potencijalno inteligentnijim nego ikada prije. Ova je knjiga poziv na istraživanje tog raskrižja: mjesto gdje se tisućljetna povijest ljudske komunikacije susreće s najnaprednjom tehnologijom 21. stoljeća. Zajedno ćemo istražiti kako se prošlost i budućnost prelamaju u sadašnjosti umjetne inteligencije, oblikujući ne samo alate koje koristimo, već možda i samu bit onoga što znači biti čovjekom koji komunicira u digitalnom dobu.

## 2 POVIJEST I EVOLUCIJA KOMUNIKACIJSKIH TEHNOLOGIJA

### 2.1 KOMUNIKACIJA KAO OSNOVA RAZVOJA CIVILIZACIJE

Svako toliko neki izum, potez pera ili tehnološko otkriće uspijeva preoblikovati naše društvo do neprepoznatljivosti. Od prvih usmenih predaja do suvremenih razgovornih agenata koji se temelje na umjetnoj inteligenciji, ljudska je potreba za komunikacijom neiscrpna pokretačka sila napretka. Svaki se civilizacijski iskorak iznova vraća na pitanje: kako možemo bolje razumjeti jedni druge i učinkovitije dijeliti informacije?

Razvoj tehnologije nije linearno, mehaničko poboljšavanje alata, već duboka promjena našeg načina razmišljanja i zajedničkog organiziranja (Diamond, 1997). Od grubih črčki u pećini, preko guščjeg pera i kaligrafije, pa sve do računala, ljudi su uvek tražili načine da smisleno prenesu ideje, emocije, iskustva i znanja. Naša potreba za komuniciranjem preko riječi – ali i daleko šireg spektra simbola, gesta i vizualnih znakova – stvorila je temelje globalno povezanog društva.

**Jezični kod** kao skup gramatike i vokabulara ovdje стоји kao osnova za složene društvene ugovore koji omogućuje suradnju, pregovaranje, razumijevanje psiholoških stanja te formiranje složenih institucija. U tom se smislu jezik pokazuje kao ključni mehanizam kojim se prenosi kolektivno sjećanje i kulturne vrijednosti, a istodobno potiču inovacije i promjene.

### 2.2 USMENA PREDAJA I ŽIVI ARHIVI

U najranijim fazama ljudske civilizacije, kada pisanje nije bilo rasprostranjeno ili uopće poznato, komunikacijska se dinamika uvelike oslanjala na **usmenu predaju**. Priče, mitovi, pjesme i legende – ispričane u krugu obitelji, plemenskih zajednica ili na svečanim okupljanjima – činile su temelj društvene kohezije i prijenosa znanja (Harris, 1989).

#### 2.2.1 Važnost usmene tradicije

Ritam, rima i melodija bili su ključni elementi u pamćenju i reproducirajući sadržaja, naročito kada se radilo o složenim informacijama, poput rođoslovja, popisa vladara ili vjerskih obreda. Djelomično iz tog razloga, pjesme i stihovi često su preuzimali ulogu svojevrsnih zapisa u društvenima bez pisma (Ong, 1982). U mnogim su kulturama pripovjedači ili **bardi** uživali visok društveni ugled jer se na njihove riječi oslanjalo cijelokupno kolektivno sjećanje zajednice. Oni su, u doslovnom smislu, bili **živi arhivi**: svojim pamćenjem i izvedbama mogli su prenijeti vijekove povijesti i kulturnih normi budućim naraštajima.

Iako na prvi pogled takav usmeni sustav može djelovati krhko, istraživanja pokazuju da je upravo **performativna narav** usmene kulture – uključujući dramske elemente, glazbu, pokrete ili obrede – omogućavala jače emotivno i kognitivno usidrenje informacija (Goody, 1987). Dugotrajna društvena praksa učenja napamet i recitiranja naučenoga “od usta do usta” sprečavala je gubitak bitnih čvorova usmene tradicije.

Također bi se moglo pomisliti da je usmeni prijenos inherentno podložan kvarenju ili iskrivljavanju informacija tijekom vremena – poput igre pokvarenog telefona na generacijskoj skali. Međutim, povijest nudi zapanjujuće primjere iznimno sofisticiranih i visoko vjernih sustava usmene predaje. Možda najimpresivniji primjer dolazi iz drevne Indije i vedske kulture. Vede, golemi korpus svetih tekstova (himni, filozofskih rasprava, ritualnih uputa) sastavljenih na arhaičnom sanskrtu, prenošene su isključivo usmeno tijekom više od tisuću godina prije nego što su konačno zapisane (Staal, 1986; Witzel, 2003). Očuvanje ovog ogromnog i kompleksnog

**Commented [BP2]:** 1. Komunikacija kao osnova razvoja civilizacije

#### Temeljna tema:

Komunikacija, temeljena na jeziku, pokretačka je snaga razvoja civilizacije. Ova cjelina pruža pregled jezika kao alata za povezivanje ljudi i tehnologije kao produžetka komunikacijskih sposobnosti.

#### Podteme:

##### 1. Jezik kao alat komunikacije i suradnje:

- Simulacija interakcije, identiteta i stvarnosti kroz jezik.
- Struktura jezika: fonološki, sintaktički i semantički obrazci.
- Prenošenje ideja i znanja među ljudima.

##### 2. Evolucija tehnologije kroz komunikaciju:

- Od mehaničkih sustava do umjetne inteligencije.
- Tehnologija kao produžetak ljudskih sposobnosti za razmjenu informacija.
- Kako je komunikacija omogućila društveni i tehnološki napredak.

##### 3. Jezik i društveni napredak:

- Kako jezik omogućuje kolaboraciju i inovaciju.
- Primjeri: pisani jezik, mediji, internet.

#### Povezivanje:

Ovo poglavlje uvodi temeljnu ideju da su jezik i tehnologija nerazdvojni u razvoju civilizacije, pripremajući teren za raspravu o procesorskoj snazi i umjetnoj inteligenciji.

**Commented [BP3]:** \*\*Razvoj komunikacijskih agenata korištenjem velikih jezičnih modela\*\* \*\*Benedikt Perak\*\*

## 1. Komunikacija kao osnova razvoja civilizacije  
Suvremena društva, koja su neprestano povezana digitalnim mrežama, duguju svoj napredak upravo toj neiscrpanoj ljudskoj težnji za razumijevanjem i boljom komunikacijom. Komunikacija nije puk prijenos podataka – ona je temeljna funkcija koja omogućava zbiljavanje ljudi, stvaranje zajednica i oblikovanje civilizacija. Upravo zato, razvoj tehnologija komunikacije i korištenje jezika kao sredstva razmjene ideja i osjetljivih iskustava predstavlja ključnu prekretnicu u povijesti ljudske civilizacije. Svaki tehnološki korak prema naprijed omogućio je društvu da efikasnije dijeli informacije, nastale ideje i vlastite osjećaje. Pogledajmo, zato, kako je razvoj tehnologije komunikacije u prošlosti, počevši od najranijih oblika usmene predaje, došao do trenutnih inovacija i području velikih jezičnih modela i komunikacijskih agenata. ## Evolucija tehnologije i jezični kod Prvi koraci ljudske civilizacije nisu zahtijevali sofisticirane uređaje, već samo riječ i gestu koja je nosila značenje. Međutim, kako su civilizacije razvijavale, javila se potreba za metodama i alatima pomoći kojih je bilo moguće pohraniti informacije, prenijeti ideje i ostvariti dalekosežne veze. Svoje korištene od najranijih oblika pisma nukleirali su u samo nekoliko jednostavnih simbola, oni koji su predstavljali društvene norme i vrijednosti, ali i nužnost jačanja institucijskih novina. Jezični kod, koji je s vremenom prešao u složnije forme, dakle nije stvorio samo osnovni temelj za snalaženje u društvenom životu, nego je izvršno direktirao norme i metodologije stvaranja civilizacijske boljke (Diamond 1997). Niti jedno novo društvo koje je naišlo na jezike automatizirali je samo da bi nastao niz složenih medusmjernih kreativizima for 9 seconds

korpusa s gotovo savršenom vjernošću (verbatim fidelity) zahtjevalo je razvoj nevjerljivojatno rigoroznih metoda učenja i recitiranja.

Ključnu ulogu u ovom procesu imali su sanskrtski žreci (brahmani), koji su prolazili dugogodišnju i izuzetno zahtjevnu obuku od najranije dobi. Njihov zadatak nije bio samo zapamtiti tekstove, već ih internalizirati s apsolutnom fonetskom preciznošću, uključujući točan izgovor svake riječi, suptilne razlike u dužini samoglasnika i, ključno, precizne tonske akcente (skr. svara), jer se vjerovalo da i najmanja promjena u zvuku može umanjiti ili poništiti svetu moć i ritualnu djelotvornost mantri (Filliozat, 2004; Staal, 1986). Kako bi osigurali ovu nevjerljivojatnu razinu točnosti, razvili su čitav niz sofisticiranih mnemotehničkih tehnika i stilova recitiranja (pāṭha):

- Samhita-pāṭha: Kontinuirano recitiranje teksta u normalnum tijeku.
- Pada-pāṭha: Recitiranje riječ po riječ, rastavljajući sve složenice i sandhi (fonetske promjene na spojevima riječi) kako bi se osiguralo točno poznavanje svake individualne riječi.
- Krama-pāṭha: Recitiranje u "koracima" od po dvije riječi (npr. riječ 1-2, 2-3, 3-4...), koje ulančano jača pamćenje veza između riječi.
- Jaṭā-pāṭha: Još složeniji uzorak (npr. 1-2, 2-1, 1-2; 2-3, 3-2, 2-3...), koji dodatno „utkiva“ redoslijed i veze.
- Ghana-pāṭha: Izuzetno kompleksan i gust način recitiranja koji uključuje još složenije permutacije i ponavljanja, koji služi kao ultimativna provjera točnosti pamćenja.

Ovi različiti stilovi recitiranja, često popraćeni specifičnim pokretima ruku (skr. mudre), služili su kao mnemonička pomagala i mehanizmi za provjeru grešaka (Rubin, 1995). Usposrednom različitim pāṭha, bilo kakva odstupanja mogla su se lako identificirati i ispraviti. Ovaj sustav pretvorio je brahmanske škole (skr. gurukule) u prave institucije za očuvanje znanja, a same brahmane u ljudske magnetofone ili, preciznije, visoko specijalizirane procesore i čuvare svetog zvuka.

Primjer vedске tradicije pokazuje da usmena kultura nije nužno bila manje razvijena od pismene; ona je jednostavno razvila drugačije, ali jednakmoće tehnologije za upravljanje informacijama, optimizirane za auditivno pamćenje i performativnu reprodukciju. Dugotrajna društvena praksa rigoroznog učenja napamet i ritualiziranog recitiranja od usta do usta, kao što vidimo u vedskom primjeru, mogla je osigurati izvanrednu stabilnost i spriječiti gubitak čak i najsigurnijih detalja usmene tradicije kroz stoljeća, demonstrirajući nevjerljivojatnu moć i otpornost ljudskog pamćenja kada je ono disciplinirano i društveno podržano.

### 2.2.2 Kolektivna imaginacija i identitet

Usmena predaja ujedno gradi **kolektivnu imaginaciju** i oblikuje identitet zajednice. Očuvane priče o mitskim herojima i zajedničkim podvizima novim su naraštajima pružale **viziju identiteta** o tome tko su oni, odakle potječu i koje su vrijednosti njihove zajednice (Levy, 1999). Čak i u današnje doba, kada je pismenost gotovo univerzalna, mnoge kulture zadržale su dijelove usmene prakse – kroz rituale, legende, izreke, folklor – kako bi očuvali emocionalnu i povijesnu vezu s prošlošću.

U život izvedbi pripovijedanja krije se i snažan element interakcije, sa starješinama, bardima ili mudracima zajednice koji često ulaze u razgovor sa slušačima, odgovarajući na pitanja i prilagođavajući svoje priče kontekstu skupine. Ovakav **dijaloški aspekt** usmene predaje, gdje

je svaka izvedba donekle jedinstvena doprinio je nadograđivanju sadržaja, a ponekad i selektivnom brisanju ili ispravljanju ranijih elemenata (Finnegan, 1970).

### 2.2.3 Društvene funkcije usmene tradicije

Društvene funkcije usmene tradicije svakako nadilaze sam čin prenošenj informacija: kroz priče, legende i mitove oblikovale su se moralne smjernice, odgajale zajednice i očuvalo zajedničko sjećanje na važne događaje. Mnoge mitologije, primjerice, prikazuju bogove koji nagrađuju pravedne i kažnjavaju nepravedne, nudeći jasnu moralnu orientaciju (Harries, 2003). Usmena se tradicija koristila i pri svečanom obilježavanju prijelaza iz jedne životne faze u drugu, pa su pjesme i stihovi prenosili sjećanja na ratne uspjehe ili uspjehe u lov i poljoprivredi (Harris, 1989).

U mnogim ranim civilizacijama bez formalnih škola upravo su ovakvi narativni obrasci djeci i mlađima pružali temeljna znanja o flori, fauni, navigaciji i gradnji (Vansina, 1985). Usmena predaja tako se pokazala stabilnijom i trajnijom no što bismo očekivali: polineziske priče o putovanjima morem i staroafrički epovi poput "Sundiata" zadržali su kroz stoljeća vrijedne povijesne tragove (Ong, 1982; Niane, 1960). Visok društveni ugled pripovjedača i stroge sankcije za one koji bi se odveć udaljili" izvornog sadržaja čuvali su jezgre narativa, omogućujući tek manju dozu kreativne prilagodbe tijekom izvedbe.

## 2.3 PISMENOST KAO PROZOR U SLOŽENE SISTEME ZNANJA

Kako su prve državne tvorevine počele rasti i uspostavljati složenije društvene, gospodarske i političke institucionalne strukture, postajalo je očito da je usmena predaja sama po sebi nedostatna za pohranu i prijenos velike količine podataka (Kramer, 1963). Zabilježiti **trgovinske transakcije, porezne obveze ili zakonske propise** na stabilan način postala je ključna potreba, što je izravno potaknulo nastanak prvih sustava pisanja. Time je rođen koncept pismenosti koji je, uz posve novi sloj društvene organizacije, nudio i novu eru u razvoju složenih sistema znanja.

### 2.3.1 Različite staze do pisma

Sumerani, Egipćani i Kinezi među prvima su pronašli načine da zvukove govora i apstraktne koncepte pretoče u piktografske, ideografske ili klinaste znakove (Ong, 1982). U sumerskoj Mezopotamiji, klinasto pismo na glinenim pločicama započelo je kao jednostavan način bilježenja poljoprivrednih viškova i trgovinskih transakcija, no postupno je obuhvatilo i mitove, zakone i epove poput čuvenog Epa o Gilgamešu". Egipatski sustav hijeroglifa, ispunjen simbolima za religijske i državne ceremonije, može se vidjeti u oslikanim i uklesanim zapisima na zidovima grobnica i hramova, gdje su služili i kao alat za političku propagandu (Parkinson & Quirke, 1995). U Kini su se razvili znakovi koji su premostili brojne dijalekte i omogućili zajedničku pismenu komunikaciju, olakšavši administraciju prostranih područja i integraciju mnogih etničkih skupina (Boltz, 1994).

Uvođenje pisma je produljilo pamćenje zajednica izvan ograničenja jedne generacije ili jedne izvedbe, ali i otvorilo mogućnost pristupanja pohranjenom znanju neovisno o tjelesnoj prisutnosti pripovjedača (Goody, 1987). Tako i nastaju prvi arhivi u kojima čuvaju, kopiraju ili prenose zapisi. U antičkom svijetu i srednjem vijeku, knjižnice poput one u Aleksandriji postaju središta prikupljanja i klasificiranja znanstvenih i filozofskih rasprava, što je stvorilo plodno tlo za napredak filozofije, matematike i raznih drugih disciplina (Pagels, 1979). U europskom srednjem vijeku, pak, samostani su zauzeli ulogu utvrda pismenosti" (Bernard, 1993). Mnogi su monasi, nerijetko tijekom cijelog života, metodično prepisivali antičke, ali i suvremene spise,

sačuvavši na taj način kontinuitet europske pisane baštine u nemirnim vremenima ratova i političke nestabilnosti. Pisana riječ je postala temeljem za razvoj složenijih oblika intelektualnog djelovanja i razmjenu ideja kroz povijest.

### 2.3.2 Društvena elita i sporost širenja

Pismenost je, ipak, dobrom dijelom prošlosti ostala ograničena na uski krug društvene elite: svećenike, znanstvenike, vladare i nekolicinu visokobrazovanih pojedinaca (Coulmas, 2003). Učenje vještine pisanja podrazumijevalo je dobar pristup rijetkim i skupocjenim materijalima poput papira i pergamenta, a službene institucije nerijetko su nadzirale proizvodnju i tumačenje pisanih dokumenata, učvršćujući tako politički i simbolički autoritet onih koji su imali moć čitati i pisati (Clanchy, 1993). Zbog sporosti prepisivanja knjiga i ograničenih ljudskih resursa za taj posao, samo je manjina imala izravan pristup rukopisima ili knjižnicama, dok je većina pučanstva ostajala isključena iz tog oblika luksuzne” sposobnosti.

U srednjovjekovnoj Europi, osobito nakon sloma Zapadnog Rimskog Carstva, samostani su funkcionalnici kao utočišta pismenosti (Bernard, 1993). Ondje su monasi, posebno u benediktinskim i sličnim redovima, prepoznali vitalnu potrebu očuvanja intelektualnog nasljeđa koje je obuhvaćalo biblijske, teološke, ali i filozofske i znanstvene tekstove antičkih civilizacija. Redovnici su u skriptorijima uporno prepisivali i iluminirali knjige, ukrašavajući marge inicijalima i kratkim komentarima, ponekad ostavljajući trag polemičkih bilježaka ili ranih znanstvenih upitnika koji su išli i protiv izvornoga teksta. Sporo, ali sustavno kopiranje tako je spasilo neka od temeljnih djela, uključujući Ciceronove rasprave, koje bi bez te brižljive prakse vjerojatno nestale pod udarima ratova i političkih nestabilnosti.

Povjesno gledano, pisani dokumenti su se koristili za rješavanje svakodnevnih administrativnih ili pravnih izazova, poput upisa poreznih obveznika, definicije imovinskih granica ili kodifikacije vjerskih propisa, ali i kao ključni katalizator budućih znanstvenih i kulturnih iskoraka. Porastom potražnje za znanjem, te razvojem tehnologije za izradu papira, pismenost je polagano napuštala isključivu domenu elite te se sve više uvlačila u šиру populaciju. Proces se ubrzao izumom tiskarskog stroja u 15. stoljeću, kada je masovno umnožavanje knjiga prešlo iz domenā samostanskih skriptorija u široku društvenu sferu, najavljujući istinsku revoluciju širenja znanja.

Upravo se u toj dobitnoj kombinaciji elitnog iskustva pismenosti i postupnog otvaranja prema široj javnosti krije podloga za daljnje komunikacijske revolucije. Pisana riječ omogućila je uspostavljanje sve složenijih zakonika, obrazovnih sustava i institucionalnih struktura koje su zahvatile gotovo sve razine života: državnu upravu, vjerske institucije, a kasnije i znanstvene i kulturne krugove. U sljedećim razdobljima, pismenost će se razvijati od rukopisnih formi do tiskanih izdanja, sve do suvremenih elektroničkih i digitalnih formata u kojima pohrana i širenje znanja dobivaju posve nove oblike i globalne razmjere.

## 2.4 TISKARSKI STROJ: PREKRETNICA U ŠIRENJU IDEJA

Povijest ljudske civilizacije obiluje prijelomnim trenucima koji su doveli do golemih skokova u načinu poimanja svijeta, organizaciji društva ili razmjeni informacija. Jedan od najvažnijih takvih trenutaka svakako je izum i masovna primjena **tiskarskog stroja** u 15. stoljeću, što se nerijetko promatra kao začetno sjeme modernog informacijskog doba (Eisenstein, 1979). Dotadašnje metode umnožavanja pisanih materijala, ponajprije ručno prepisivanje u skriptorijima, bile su spore, podložne pogreškama i vrlo ograničenog dosegaa. Korištenjem Gutenbergova postupka mehaničkog otiskivanja, knjige su mogle istodobno biti tiskane u višestrukim primjercima, širiti se na različita geografska područja i, u konačnici, stići do znatno veće publike.

#### **2.4.1 Masovna proizvodnja i demokratizacija” znanja**

Gutenbergov izum dramatično je skratio vrijeme koje je bilo potrebno za izradu jedne knjige. Uz to, cijena tiskanih materijala počela je padati, čineći ih dostupnijima širem krugu ljudi (Febvre & Martin, 1976). Širenje tiskanog materijala ubrzo je dovelo do višestrukih promjena koje su postavile temelje dalnjim civilizacijskim iskoracima. Prije svega, omogućilo se ubrzanje i dostupnije protjecanje ideja: znanstvenici, filozofi i drugi mislioci napokon su mogli tiskati vlastite radove u znatnim nakladama, pa je provjera i usporedba rezultata sličnih istraživanja postala znatno jednostavnija. Primjerice, matematičari i astronomi, čija bi se otkrića prije prenosila polako i u malobrojnim rukopisnim verzijama, sada su svoje radove mogli proširiti diljem Europe, čime su se istraživanja različitih skupina mogla međusobno brže uspoređivati i dopunjavati.

U takvom se okruženju probudila i nova uloga sveučilišta: ona nisu više bila tek mjesto gdje su profesori i studenti razmjenjivali ideje u ograničenoj učionici”, nego su, prepoznavši važnost tiskane knjige, počela graditi goleme sustave pohrane – rane biblioteke. Zahvaljujući organiziranom arhiviranju i klasificiranju raznovrsnih tiskanih izdanja, sveučilišta su se preobrazila u sjedišta za kritička i interdisciplinarna istraživanja, stvarajući infrastrukturu znanja” koja je dodatno ubrzavala znanstvenu razmjenu. Upravo su na tim temeljima postupno nastali znanstveni časopisi i druge oblike skupnih publikacija, a s vremenom je sve postalo još dinamičnije kako se komunikacija ubrzala i standardizirala.

No, možda je najizrazitija posljedica sve veće dostupnosti knjiga bila postepena demokratizacija” znanja (Anderson, 1983). Pismenost i pristup tiskanim materijalima više nisu bili strogo ograničeni na klerikalnu ili aristokratsku elitu; iako je taj proces tekao polako, sve više pojedinaca moglo je prodrijeti u političke, znanstvene i vjerske rasprave, nakupljujući kritičko znanje i time često potresajući stare društvene hijerarhije. Novo otvoreno” polje učenja, barem u teoriji, omogućilo je ambicioznima i radoznalima da iz grade vlastitu prosvijetljenu” poziciju, što je produbilo i razgranalo misaone tokove čitavog razdoblja.

#### **2.4.2 Reformacija i znanstvena revolucija**

Tiskarski je stroj snažno utjecao na religijska previranja i reformacijske pokrete, poput onog koji je pokrenuo Martin Luther (1517. i dalje). Lutherove teze tiskane su i umnožene diljem Europe, uznenirivi crkvenu hijerarhiju koja se dotad oslanjala na kontrolu pristupa vjerskim tekstovima. Nadalje, znanstveni radovi – osobito tijekom 17. stoljeća – cirkulirali su kroz tiskane **disertacije” i traktate**”, ubrzavajući **znanstvenu revoluciju** (Febvre & Martin, 1976). Mnoge ideje, koje su prije ostajale ograničene na malobrojne erudite, postale su šire dostupne i time potaknule lančanu reakciju inovacija.

#### **2.4.3 Globalno širenje i kulturni utjecaj**

Posrijedi nije bio samo europski fenomen. Kako su se tiskare širile i u druge dijelove svijeta, knjige su postale **putujući svjedoci”** kulturnih praksi i filozofija, nerijetko spajajući različite civilizacije. Primjerice, u islamskom svijetu, početni otpor prema tiskanim Kur'anima (zbog bojazni da bi se sveta riječ mogla krivo reproducirati) s vremenom je ustuknuo pred praktičnim prednostima tiska, čime je došlo do bržeg prijenosa religijskih komentara i znanstvenih tekstova (Roper, 2017). Dok je u Kini drvorezno tiskanje bilo poznato stoljećima prije Guttenberga, (Needham, 1985) no tek se u 19. stoljeću, europski model tiskarstva počeo se vraćati u Kinu i utjecati na modernizaciju kineskih tiskarskih pogona, a posredno i na kineski izdavački sektor.

#### **2.4.4 Preteča informacijskog društva**

Upravo zahvaljujući tiskarskom stroju stvorena je **kultura knjige** kao trajnog artefakta. Taj se izum” čini ključnim za uvođenje koncepta **masovne pismenosti**, koji će kasnije, s dalnjim tehnološkim razvojem (poput tiskanja novina, časopisa i, konačno, digitalnih izdanja), dosegnuti dosad neviđene razmjere.

Neki povjesničari i sociolozi (McLuhan, 1964; Eisenstein, 1979) tvrde da je upravo masovni tisak odgojio” europeizirano stanovništvo na kulturi čitanja i interpretacije, postavljajući temelje za koncepte javnog mnijenja, nacionalnih identiteta (Anderson, 1983) i pluralističkih rasprava. Bez tog prethodnog koda razmjene i razumijevanja pisanih informacija, kasnije pojave, poput radija, televizije ili interneta, ne bi imale jednaku plodno tlo za ubrzalu recepciju.

#### **2.4.5 Nasljeđe za daljnji razvoj komunikacije**

Gutenbergova revolucija ostavila je golem trag u svjetskoj kulturnoj i intelektualnoj povijesti. Tiskanjem **Biblike, znanstvenih rasprava, literarnih djela** i brojnih drugih tekstova, društvo je postalo tekstualno povezanije, što se, naravno, odrazilo punim potencijalom u narednim stoljećima, kako se širila pismenost i tehnološke inovacije.

U kasnijim etapama razvoja komunikacijskih tehnologija, tradicija masovne proizvodnje pisane riječi bila je ključna za pojavu novina i časopisa, što je pak izravno utjecalo na političku transparentnost i društvenu mobilizaciju. Daljnji su tehnološki prodori (poput stroja za linotip ili offsetnog tiska) i pojeftinjenja papira još više ubrzali proizvodnju pisanih medija, stvarajući uvjete za ono što ćemo danas nazvati ekosustavom informacija, unutar kojeg je zatim rođen i **digitalni svijet**.

### **2.5 TELEGRAF I TELEFON: POČETAK ELEKTRONIČKE KOMUNIKACIJE**

Napredak koji je donijelo 19. stoljeće u području komunikacijskih tehnologija bio je toliko snažan da je iz temelja promijenio dotadašnje poimanje prostora i vremena. Dok je ranije prijenos vijesti i ideja ovisio o brzini konja, parobroda ili kočije, s uvođenjem **telegraфа** i, ubrzo zatim, **telefона**, mogućnost gotovo trenutnog povezivanja kontinenata postala je nova normalnost (Standage, 1998). Na taj je način rođena elektronička komunikacija i započeli dotad nevideni razmjeri procesa globalizacije.

#### **2.5.1 Telegraf: nove brzine i novi horizonti**

Jedan od središnjih aktera ove priče jest **Samuel Morse**, američki slikar i izumitelj, koji je tijekom 1830-ih i 1840-ih godina razvio prvi praktični električni telegraf temeljen na tankim žicama i vlastitom kodu točaka i crtica (Morseov kod). Već 1844. uspostavljena je prva komercijalna telegrafska linija između Washingtona D.C. i Baltimorea, a prva poslana poruka – “What hath God wrought?” – simbolizirala je svu težinu i uzbudjenje tog tehnološkog čuda” (Coe, 1993).

Telegrafska je inovacija omogućila da se poruke, koje bi inače putovale danima ili tjednima, prenesu doslovno u minutama (Standage, 1998). Ova **munjevita** promjena posebno je dobro došla u sferama poslovanja, politike i novinarstva. Primjerice, Associated Press koristio je telegrafske linije kako bi ostvario skoro trenutačnu razmjenu vijesti u zemlji, a finansijska tržišta diljem SAD-a prilagođavala su se u hodu novim cijenama i informacijama o tržišnim kretanjima (Blondheim, 1994).

Tijekom Američkog građanskog rata (1861.–1865.), zapovjednici su koordinirali trupe na različitim bojištima uz pomoć telegrafa, što je omogućilo **brže i točnije** donošenje odluka

(Wheeler, 2000). U kasnijim desetljećima, prijenos ratnih izvještaja i diplomacija ovisili su o telegrafskim linijama, često preko podmorskih kabela koji su spajali kontinente.

Širenje telegrafskih mreža potaknulo je i **standardizaciju vremena** te uvođenje vremenskih zona, posebice radi koordinacije željezničkog prometa i međunarodnih telegrafskih poruka (Bartky, 2007). Bio je to prvi ozbiljan korak prema shvaćanju svijeta kao međusobno ovisnog sustava”, gdje su precizne informacije o točnom vremenu bile ključne za gospodarstvo i logistiku.

### 2.5.2 Telefon: intimna revolucija na žici”

Nešto osobniji, ali ne manje spektakularan pomak donio je **Alexander Graham Bell**, patentiravši 1876. godine telefon – napravu koja je prenosila zvučne impulse i intonacije ljudskog glasa izvan dotadašnjih granica (Bell, 1876).

Za razliku od telegrafskih točaka i crtica, telefon je ljudima omogućio da se čuju” kao da su u istoj sobi, bez potrebe za ikakvim kodom. Brzi poslovni dogovori, obiteljski razgovori preko oceana ili međunarodna koordinacija postali su realnost (Fischer, 1992). Upravo taj osjećaj **neposrednosti i intimnosti** promjenio je svakodnevni život: ljudi su se lakše snalazili u kriznim situacijama, organizirali događaje i održavali veze na daljinu.

Kako su telefonske linije polako osvajale velike gradove, a potom i ruralne sredine, nastajala je globalna mreža koja je svijet transformirala u *zametke globalnog sela*” (Carey, 1989). Ljudi su, po prvi put, uspjeli dogovoriti sastanke i transakcije u realnom vremenu bez posredništva poslužitelja ili telegrafskih operatera.

Povećana brzina razmjene informacija i jednostavnost govornog kontakta imali su izravan utjecaj na razvoj gospodarstva. Tvrtke su brže donosile odluke i mogle su reagirati na tržišne promjene i u udaljenim regijama. Širenje telefonskih usluga postalo je važan segment gospodarskog rasta i otvaranja novih radnih mjesta, stvarajući inovacije u inženjerstvu, tehnologiji i uslužnim djelatnostima (Chandler & Cortada, 2000; Gleick, 2011). U političkoj i društvenoj sferi, mogućnost da se informacije i stavovi brzo dijele doprinijela je **demokratizaciji** i ubrzanim protoku ideja, ponajprije u organiziranju političkih kampanja i društvenih pokreta (Standage, 1998).

### 2.5.3 Preoblikovanje društvene dinamike

**Telegraf i telefon** nisu samo ubrzali komunikaciju već izravno utjecali na to kako su se ljudi počeli odnositi prema prostoru i vremenu. Međusobni razgovori između gradova ili kontinenata prestali su biti ekscentrična iznimka i postali redovita praksa, skraćujući” fizičke udaljenosti. Više nije bilo nužno čekati tjednima ili mjesecima na pismo, a trenutna” veza s udaljenim kontaktnim točkama postala je paradigmom nove komunikacijske kulture.

Uvođenje **vremenskih zona** i potrebe globalne sinkronizacije (Bartky, 2007) ne bi bilo moguće bez brzog i pouzdanog prijenosa informacija preko telegrafskih i telefonskih mreža. Industrije poput željeznica, brodarstva i financija ovisile su o točnim vremenima polazaka i dolazaka, te su se morale uskladivati sa sličnim sustavima diljem svijeta.

Početak globalnog sela” također je značio korake prema stvaranju svjetske telekomunikacijske mreže, koju su gradili gigantski poslovni koncerni, često uz podršku državnih strategija (Lipartito, 2003). Ta je infrastruktura, koja je svojedobno uključivala podmorske kabele i goleme telefonske centrale, postala temelj za kasniji razvoj interneta i drugih digitalnih tehnologija.

#### **2.5.4 Impuls za buduće komunikacijske revolucije**

Gledajući unazad, možemo reći da je 19. stoljeće kroz telegraf i telefon pripremilo teren” za **kasnije pojave radija, televizije, a zatim i interneta**. Uvođenje elektroničke komunikacije oblikovalo je shvaćanje da je brzina protoka informacija postala gospodarska, politička i društvena konkurentska prednost (Chandler & Cortada, 2000). Ujedno je stvorena nova društvena dinamika u kojoj **intimnost razgovora**, čak i na udaljenosti od tisuća kilometara, postaje normalizirana.

Upravo će se na ovoj infrastrukturi — zbijenoj mreži žica i telefonskih centrala — kasnije graditi globalna mreža optičkih vlakana, satelitskih veza i bežičnih mreža, koje će dodatno razbiti prostorne i vremenske zapreke. Dok se povijest komunikacije postupno približava masovnim medijima 20. stoljeća, postaje očito da su **telegraf** i **telefon** bili ključni uvod u elektroničku eru u kojoj živimo i danas.

### **2.6 RADIO I TELEVIZIJA KAO KATALIZATORI MASOVNOG ISKUSTVA**

U prvoj polovici 20. stoljeća, dok je svijet još bio očaran mogućnostima telegrafa i telefona, stupili su na scenu masovni elektronički mediji, od kojih se posebno ističu radio i televizija. Ovi su se mediji pokazali ne samo kao tehnološke inovacije, već i kao snažni pokretači kulturne homogenizacije, političkog oblikovanja i društvenog zajedništva (McLuhan, 1964). Priča o njihovu razvoju povezuje vizionarstvo, geopolitičke kontekste, komercijalne interese i kreativne potvhvate — a sve na tragu ideje da je komunikacija jedan od temeljnih izvora civilizacijskog razvoja.

#### **2.6.1 Uspon radija i glas” zajedničke stvarnosti**

**Radio** je iznikao iz ideja i eksperimentalnih pothvata krajem 19. stoljeća, no tek su pioniri poput Guglielma Marconija (Marconi, 1902) i Nikole Tesle osigurali praktične temelje bežične komunikacije. Rane radiopostaje često su bile usmjerene na prijenos vijesti i glazbe, no s vremenom su pridobile i širu ulogu: postale su glavni izvor ratnih izvještaja, sportskih prijenosa i zabavnih emisija. U godinama koje su prethodile Drugom svjetskom ratu, radio je bio ključno oruđe za propagandu i mobilizaciju javnosti. Primjerice, nacistička Njemačka intenzivno je koristila Volksempfänger” prijemnike kako bi širila državni narativ, dok su saveznici uz pomoć radijskih servisa poput BBC-ja doprli do okupiranih područja (Briggs & Burke, 2009).

Jedan od najslavnijih primjera moći radija kao komunikacijskog sredstva jest emisija “**War of the Worlds**” Orsona Wellesa 1938. godine. Iako je većina slušatelja bila svjesna da se radi o adaptaciji znanstveno-fantastične priče, dio publike doživio je paniku misleći da je riječ o stvarnoj invaziji Marsovaca (Cantril, 1940). Taj događaj pokazuje kako je, zahvaljujući napetom i uvjerljivom radijskom formatu, bilo moguće utjecati na kolektivnu svijest brže i snažnije nego ikada prije. Radio je tako postao **glas” zajedničke stvarnosti**, stvorivši dotad neviđenu sponu među ljudima koji su, u istom trenutku, mogli čuti jedinstven sadržaj, osjećajući se kao dio velike, integrirane zajednice (Lewis, 1991).

#### **2.6.2 Televizija i vizualna dimenzija zajedničke imaginacije**

Dok je radio transformirao slušnu dimenziju javne sfere, **televizija** je u sredini 20. stoljeća (posebice nakon Drugoga svjetskog rata) učinila isto s vizualnom kulturom. Rani eksperimenti Johna Logiea Bairda 1920-ih godina (Burns, 2000) te kasnije komercijalne inicijative u Sjedinjenim Državama, Ujedinjenom Kraljevstvu i drugim zemljama, vodili su k tome da se televizijski prijemnici rasprostrane diljem svijeta. Već tijekom 1950-ih, televizija je postala

pristupačna široj populaciji, a obitelji su okupljanje uz mali ekran” doživljavale kao društveni ritual (Spigel, 1992).

Televizija je unijela **novu razinu emocionalnog angažmana**: gledatelji su mogli vidjeti – ne samo čuti – političke vođe, sportske legende, glazbene izvođače i filmske zvijezde. Time je, primjerice, predsjednička debata između Johna F. Kennedyja i Richarda Nixon-a 1960. (Kraus, 1960) ostala zapamćena kao jedan od prvih globalnih testova moći televizijskog dojma: analitičari su primijetili da je vizualna karizma imala gotovo jednak značenje kao i retorička vještina.

Pored političkog utjecaja, televizija je značajno oblikovala **popularnu kulturu**, unosiša masovnu zabavu u dnevne boravke i istovremeno nametala određene stilove života, mode i vrijednosti (Williams, 1974). Emisije poput *I Love Lucy*” (1950-e) ili kasnije hit-serije poput *Dallas*” (1980-e) postale su globalne pop-fenomenom, predstavljajući američki način života i posredno potičući međunarodni kulturni dijalog — ali i određene stereotipizacije.

#### 2.6.3 Masovna komunikacija i homogenizacija kulture

Doba radija i televizije stvorilo je posebnu vrstu **zajedničkog kulturnog iskustva**: ljudi iz različitih geografskih i društvenih miljeva gledali su ili slušali iste sadržaje, gotovo u isto vrijeme. Taj je fenomen McLuhan (1964) opisao kao globalno selo”, referirajući se na ideju da elektronički mediji zblžavaju zajednice nalik na tradicionalno selo, gdje se svatko bavi zajedničkim vijestima i raspravama. Istovremeno, širenje istih informacija i vrijednosti dovelo je do određene **kulturalne homogenizacije**.

Na primjer, mladi iz raznih dijelova Europe ili Azije mogli su zajedno gledati dodjele nagrada kao što su *Oscar* ili *Eurovision Song Contest*, stječući dojam da su dio iste svjetske” zabavne scene. To je ujedno potaknulo i unifikaciju potrošačkih navika: globalne reklamne kampanje na televiziji povećale su prodaju brendiranih proizvoda (Coca-Cola, Levi's, McDonald's), oblikujući ukuse i želje potrošača na nadnacionalnoj razini (Schudson, 1993).

#### 2.6.4 Politička sfera i elektroničko javno mnjenje

**Radio i televizija** transformirali su i političku komunikaciju. Lideri su shvatili da se putem eteričnih” i vizuelnih” poruka mogu izravno obratiti biračima, zaobilazeći do tada neizbjegljive novinske posrednike (Brants & Voltmer, 2011). Primjerice, Franklin D. Roosevelt koristio je “fireside chats” (godišnje radijske obraćanje naciji) kako bi stanovnike SAD-a informirao o ekonomskoj situaciji i ratnim izazovima. Birači su, slušajući predsjednika u intimi vlastite dnevne sobe, stjecali osjećaj neposrednog kontakta koji se ranije nije mogao ostvariti.

Ne samo da su se time stvorile nove mogućnosti za političare, već i za **manipulacije**. Režimi s autoritarnim tendencijama brzo su uvidjeli prednost kontroliranog prijenosa” informacija. Državne televizije postale su prvi izvor istine” za široke slojeve stanovništva, jer je tek dio publike imao pristup alternativnim izvorima. U otvorenijim društвima, pak, televizijske političke debate postale su odlučujuće za ishod izbora, a način na koji netko izgleda i govori pred kamerom mogao je imati gotovo jednak utjecaj kao i snaga argumenata (Kraus, 1960).

#### 2.6.5 Komunikacija kao temelj civilizacije: pogled unaprijed

Promatrajući utjecaj radija i televizije, jasno je da su masovni mediji iz korijena redefinirali društvene odnose, **kulturne obrasce i političke procese**. Kao ključne faze u povijesti komunikacije, radio i televizija prikazuju kakvu moć ima tehnologija kada prodire u sferu ljudskog iskustva: oni su sinkronizirali masovne emocije, homogenizirali jezik” globalne zabave i

transformirali javno mnjenje. Na sljedećim razinama, internet i digitalni mediji preuzimaju štafetu, otvarajući nove mogućnosti participacije i personalizacije, ali i nove izazove u pogledu privatnosti, dezinformiranja i društvenih nejednakosti.

Upravo se na tim povijesnim temeljima — od prvih radiovalova do „živog“ televizijskog prijenosa — razvijaju i suvremene platforme te komunikacijski agenti. Oni s jedne strane nastavljaju ideju da je komunikacija zaslužna za organiziranje civilizacije, dok se s druge strane suočavaju s gotovo beskonačnim tehnološkim potencijalima i društvenim dilemama koje iz toga proizlaze. U nastavku ćemo sagledati kako su te nove tehnologije, potaknute internetom i umjetnom inteligencijom, nastavile i produbile procese koje su radio i televizija započeli.

Tablica 1: Brzina, količina informacija i propusnost kroz različite tehnologije komunikacije

Tehnologija	Vrijeme razvoja	Brzina komunikacije	Količina informacija	Propusnost (približno)	Domet i doseg	Ključne karakteristike i utjecaj
Govor	Prije ~100.000 godina	Trenutna (lice u lice)	Ograničena na riječi i geste	~3 kbit/s (govorni signal)	Lokalni (slušni doseg)	Osobna komunikacija, temelj društvene interakcije
Pisani jezik	Prije ~5.000 godina	Sporo (putem glasnika ili pisama)	Veća količina nego govor; trajna pohrana	~5 znakova/s (ručno pisanje)	Regionalni	Pohrana znanja, omogućuje razvoj civilizacija
Tiskarski stroj	15. stoljeće	Brže od ručnog prepisivanja	Masovna proizvodnja tekstova	~1000 stranica/dan (tiskanje)	Regionalni do globalni	Širenje pismenosti, znanstvena i kulturna revolucija
Telegraf	19. stoljeće	Brzina svjetlosti kroz žicu	Tekstualne poruke (Morseov kod)	~1–10 bit/s	Medunarodni	Trenutni prijenos na velike udaljenosti, povezuje kontinente
Radio	Početak 20. stoljeća	Brzina svjetlosti (bežično)	Audio sadržaj (glas, glazba)	~10 kHz (AM radio), ~200 kHz (FM radio)	Globalni	Masovna komunikacija, informiranje i zabava javnosti
Televizija	Sredina 20. stoljeća	Brzina svjetlosti (bežično/kabel)	Audio i video sadržaj	~6 MHz (analogni TV)	Globalni	Vizualni medij, oblikovanje javnog mnijenja, reklame
Internet	Kraj 20. stoljeća	Trenutna (brzi protok podataka)	Neograničena količina digitalnih podataka	Kilobit/s do gigabit/s	Globalni	Digitalna revolucija, interaktivnost, demokratizacija informacija
Al sustavi	Početak 21. stoljeća	Trenutna (obrada u stvarnom vremenu)	Velike količine podataka, personalizacija	Megabit/s do terabit/s (obrada podataka)	Globalni, personalizirani	Automatizacija, prilagodba, napredno razumijevanje jezika

## 2.7 INTERNET I DIGITALNA REVOLUCIJA

Negdje na razmeđi 20. i 21. stoljeća, u vremenu kada se činilo da je većina komunikacijskih izazova riješena brzim i globalnim povezivanjem putem telefona, radija i televizije, na horizontu se pojavio sasvim nov komunikacijski ekosustav: internet. U samo nekoliko desetljeća, taj je sustav, isprva zamišljen kao eksperimentalni projekt američke ARPA-e (Advanced Research Projects Agency), prerastao u temeljnu infrastrukturu za razmjenu informacija, suradnju i društvene promjene (Leiner et al., 2009).

U početku, ARPANET je zamišljen kao istraživačka mreža koja će povezivati znanstvenike i sveučilišta diljem Sjedinjenih Država. Primjera radi, Sveučilište u Kaliforniji (UCLA) i Institut za istraživanja Stanford (SRI) bili su među prvim institucijama koje su se uključile u projekt, omogućujući istraživačima bržu razmjenu podataka i koordinaciju eksperimenta. Ono što je u ranim sedamdesetima izgledalo kao ništa tehnološka ideja rezervirana za uski krug znanstvene zajednice, uskoro se pretvorilo u fundamentalni izum koji će obilježiti čitavu kasniju eru digitalne komunikacije.

Pravi skok dogodio se 1989. godine, kada je Tim Berners-Lee, znanstvenik iz CERN-a, osmislio World Wide Web (Berners-Lee, 1990). Ako je ARPANET bio geografska žila kućavica” akademskih institucija, WWW je postao lice” interneta, pretvarajući ga u platformu pristupačnu svakom korisniku. Kroz hipertextualne poveznice i jednostavne web-preglednike, ljudi su po prvi put mogli šetati” kroz daleke digitalne krajeve, istovremeno dijeleći dokumente, slike, pa čak i rane oblike multimedijalnih sadržaja.

Razvoj internetskih platformi i alata u posljednja se tri desetljeća pokazao kao središnji pokretač transformacija u načinu na koji ljudi međusobno komuniciraju, surađuju i zajednički oblikuju društvene odnose. S jedne strane, ovi procesi otvorili su nebrojene mogućnosti za umrežavanje, razmjenu ideja i stvaranje globalnih zajednica, dok su s druge strane potaknuli brojne kontroverze, prvenstveno o privatnosti i kvaliteti novonastalih odnosa.

### 2.7.1 Proširenje komunikacijskih kanala

Među prvim masovno prihvaćenim internetskim uslugama ističe se e-pošta, koja je devedesetih godina 20. stoljeća unijela radikalni pomak u svakodnevnu poslovnu i osobnu razmjenu informacija. Za razliku od tradicionalne pošte i faksa, e-pošta je omogućila trenutačnu isporuku poruka bez obzira na geografsku udaljenost (Shapiro & Varian, 1999). Primjerice, globalne tvrtke počele su povezivati svoje urede na različitim kontinentima gotovo u realnom vremenu, ubrzavajući donošenje odluka te podižući razinu koordinacije među timovima. Istodobno, privatni korisnici koristili su e-poštu za svakodnevnu komunikaciju s obitelji i prijateljima, čineći svijet manjim mjestom" uz minimalan trošak i uz neusporedivu brzinu razmjene podataka.

### 2.7.2 Suradnja i kolaborativni alati

U nadogradnji na e-poštu, pojavili su se i različiti kolaborativni sustavi poput chat-roomova, foruma te instant messaginga (npr. ICQ, AIM, MSN Messenger). Ti rani oblici digitalne razmjene ubrzo su doprinijeli razvoju inovativnih platformi za grupni rad u oblaku", poput Basecampa ili Slacka, koje su značajno pojednostavile timsku suradnju na projektima, osobito kada je riječ o razvoju softvera, organizaciji međunarodnih konferencija ili koordinaciji daljinskog učenja (Benkler, 2006). Posebno je zanimljiv slučaj Mozilla Foundationa i izrade web-preglednika Firefox, gdje se tisuće volontera iz cijelog svijeta koordiniralo putem chat-kanala i bug-tracking sustava, demonstrirajući kako se složeni projekti mogu izgraditi zahvaljujući dobro organiziranom online zajedništvu (Raymond, 1999).

### 2.7.3 Ekspanzija društvenih mreža

Nakon učvršćivanja paradigme brze digitalne komunikacije, krajem 1990-ih i početkom 2000-ih dolazi do pojave rane generacije društvenih mreža, kao što je SixDegrees.com (pokrenut 1997.). Iako inicijalno ograničenog dosega, SixDegrees je utro put kasnijim mega-platformama poput MySpacea (2003.) i Facebooka (2004.), na kojima se iscrtao zamah masovnog online druženja" (Boyd & Ellison, 2007).

Nakon inicijalnog uzleta MySpacea, koji je početkom 2000-ih postao sinonim za glazbeno otkrivanje i interakciju s bendovima, internet je nastavio iznjedravati sve sofisticiranije oblike društvenih mreža i online platformi. Iako je MySpace tijekom određenog razdoblja bio ključni digitalni prostor za promociju glazbenika, DJ-eva i mladih talenata (Boyd, 2006), njegov je format ubrzo zasjenio Facebook, koji je uveo koncept stvarnih, **pravičnijih i privatnijih** korisničkih profila (Ellison, Steinfield & Lampe, 2007). Uklanjanjem velikog dijela anonimnosti, Facebook je resetirao dinamiku digitalne socijalnosti – umjesto nadimaka" i avatara", korisnici su predstavljali sebe kao stvarne osobe sa stvarnim imenima, fotografijama i društvenim statusima.

S vremenom su društvene platforme poprimile različite oblike i orientacije. Pojavili su se servisi za razmjenu kratkih poruka ili fotografija (primjerice, **Twitter** 2006. godine i **Instagram** 2010. godine) koji su uspješno odgovorili na potrebu korisnika za što sažetijim, brzim i često vizualno atraktivnim komunikacijama (Boyd & Ellison, 2007). Zatim su tu i platforme poput **TikToka** (pokrenutog 2016. godine, ranije znanog kao Musical.ly), koje su kroz kratki video format stvorile

novi prostor za spontano, kreativno i viralno izražavanje, a istovremeno i za nepredvidivo širenje trendova, glazbe te kulturnih memova (Kaye, Chen & Zeng, 2022).

Ova je migracija offline” odnosa na digitalne prostore vođena raznovrsnim faktorima. U prvom redu, platforme su nudile (i nastavljaju nuditi) atraktivne značajke poput **dijeljenja fotografija, statusnih objava i komentara u realnom vremenu**, popraćene mogućnostima lajkanja” i tagiranja” (Wellman & Haythornthwaite, 2002). Te su značajke otvorile nove putove identitetskog izražavanja: u jednoj objavi, pojedinac može prikazati svoju profesionalnu stranu, u drugoj ležernu obiteljsku atmosferu, dok u trećoj izražava svoj politički stav.

Istodobno su se razvile i **nišne mreže** za specifične interese i ciljeve, potvrđujući tezu o raznolikim mogućnostima koje online platforme mogu pružiti. Primjerice, **LinkedIn** (pokrenut 2003.) fokusira se na profesionalne kontakte, omogućujući tvrtkama i stručnjacima da lakše “mapiraju” poslovne prilike, dok je **Couchsurfing** (2004.) postao platforma za putnike spremne na kulturnu razmjenu i iskustvo boravka kod lokalnih stanovnika bez naknade (Rosen, Lafontaine & Hendrickson, 2011). Ovi primjeri pokazali su da svaka platforma posjeduje vlastite **komunikacijske kodekse**”, koji oblikuju načine na koje se ostvaruju kontakti i gradi reputacija unutar određene online zajednice.

#### 2.7.4 Digitalna Plemena i Trgovi Znanja: Anatomija Uspješnih Online Zajednica

Dok su rane društvene mreže težile stvaranju širokih, općih digitalnih trgovaca, internetski krajolik ubrzo je počeo cvjetati u specijalizirane ekosustave – virtualne zajednice posvećene stvaranju, dijeljenju, vrednovanju i raspravi o vrlo specifičnim vrstama sadržaja ili interesa. Platforme su postale fascinantni društveni laboratoriji gdje se korisnici okupljavaju, razvijajući vlastite mehanizme upravljanja, sustave reputacije i jedinstvene mikro-kulture koje oblikuju ponašanje i interakciju unutar svojih digitalnih zidina.

Uzmimo za primjer Reddit. Ova platforma nije toliko jedinstvena zajednica koliko golema, sprawling digitalna metropola sastavljena od tisuća, ako ne i milijuna, tematskih podzajednica poznatih kao *subredditi*. Svaki subreddit funkcioniра kao zasebna četvrt ili čak minijaturna digitalna grad-država, često s vlastitim, pomno razrađenim skupom pravila, normi ponašanja i timom volonterskih moderatora koji djeluju kao neka vrsta lokalne uprave (Massanari, 2017). Spektar je nevjerojatan: od strogo moderiranih akademskih foruma poput r/science ili r/AskHistorians, gdje se zahtijevaju detaljne reference i dubinska analiza, do kaotičnih i humorističnih kutaka poput r/funny ili nišnih zajednica posvećenih nazužim mogućim interesima. Ono što Redditu daje dinamiku jest njegov temeljni mehanizam *upvote/downvote*, koji djeluje kao neumorni, kolektivno pokretan kuratorski stroj: sadržaj koji zajednica smatra vrijednim ili zanimljivim izdiže se na vrh, dok se manje relevantan ili nekvalitetan potiskuje u digitalnu pozadinu. To je neprekidni referendum o relevantnosti.

Ako je Reddit užurbani grad s bezbroj različitih četvrti, onda je Stack Overflow (dio šire Stack Exchange mreže) njegova precizno organizirana tehnička knjižnica ili možda visoko specijalizirani cehovski dom za programere. Nastao iz potrebe za brzim i točnim odgovorima na konkretna tehnička pitanja, Stack Overflow je postao gotovo nezaobilazna referentna točka u svijetu razvoja softvera (Mamykina et al., 2011). Njegov uspjeh leži u rigoroznom fokusu na kvalitetu i meritokraciju. Sustav bodovanja (reputacije) i prihvaćanja najboljih odgovora stvara snažan poticaj za pružanje točnih, jasnih i korisnih rješenja. Zajednica aktivno održava visok omjer signala i šuma, obeshrabrujući nejasna pitanja ili odgovore temeljene na mišljenju, čime se stvara izuzetno učinkovit mehanizam za kolektivno rješavanje problema i arhiviranje specifičnog tehničkog znanja.

Drugačiji evolucijski put slijedio je Discord. Izvorno pokrenut kao alat za glasovnu komunikaciju među igračima videoigara, brzo je prerastao svoju početnu nišu i postao platforma za širok spektar zajednica (Seering et al., 2019). Za razliku od asinkronih foruma poput Reddit-a ili Stack Overflow-a, Discord nudi osjećaj prisutnosti i interakcije u stvarnom vremenu, više nalikujući mreži živahnih digitalnih klubova ili društvenih centara. Njegova struktura temeljena na serverima, s kanalima organiziranim po temama (tekstualnim, glasovnim, video), te mogućnost integracije botova za automatizaciju zadataka i moderiranje, pokazala se izuzetno fleksibilnom. Danas ga koriste ne samo igrači, već i obrazovne institucije, umjetnički kolektivi, grupe za podršku, pa čak i tvrtke za internu komunikaciju i izgradnju vanjske zajednice, stvarajući prostore za fluidniju i neposredniju interakciju.

Naposljetku, ne smijemo zaboraviti ni projekte otvorenog koda (open-source) i kolaborativne repozitorije poput GitHub-a. Iako nisu društvene mreže u klasičnom smislu, platforme poput GitHub-a funkcioniraju kao ogromna digitalna gradilišta i radionice, te istovremeno kao moće komunikacijske mreže koje povezuju milijune programera i tehničkih stručnjaka diljem svijeta (Dabbish et al., 2012). Kombinacija alata za verzioniranje koda (poput Gita), sustava za praćenje problema (issue trackers), mehanizama za predlaganje i recenziranje promjena (pull requests) te integriranih foruma za raspravu stvara visoko strukturiran komunikacijski okvir optimiziran za složenu tehničku suradnju. Uspjeh projekata poput operativnog sustava Linux, web okvira React ili AI biblioteke TensorFlow svjedoči o nevjerojatnoj moći ovog modela distribuirane kolektivne inteligencije, gdje tisuće pojedinaca, često bez izravnog susreta, uspijevaju zajedno izgraditi neke od najkompleksnijih softverskih artefakata današnjice.

Ovi raznoliki primjeri – od gradova-država Reddit-a i cehova Stack Overflow-a, preko klubova Discorda do globalnih gradilišta GitHub-a – ilustriraju bogatstvo i inventivnost digitalnih oblika udruživanja. No, ovaj živredni digitalni život nije liшен sjena i izazova. Održavanje zdravlja i funkcionalnosti ovih zajednica predstavlja stalnu borbu. Moderiranje sadržaja je često Sizifov posao, borba protiv govora mržnje, uzneniravanja, dezinformacija i spama zahtijeva ogromne resurse i često dovodi do kontroverzi oko cenzure i slobode govora (Gillespie, 2018; Roberts, 2019). Problem informacijskog preopterećenja (information overload) je sveprisutan – kako pronaći relevantan signal u zaglušujućoj buci sadržaja? (Bawden & Robinson, 2009). Nadalje, algoritamski sustavi rangiranja i preporučivanja koji pokreću mnoge od ovih platformi mogu nenamjerno stvarati filter mjejhure i echo komore, ograničavajući izloženost korisnika različitim perspektivama (Pariser, 2011).

Upravo u ovom kompleksnom i često kaotičnom okruženju, komunikacijski agenti pokretani naprednom umjetnom inteligencijom počinju se pojavljivati kao potencijalno rješenje, ali i kao novi izvor izazova. Postoji rastući interes za korištenje AI alata za pomoć u moderiranju (npr. automatsko označavanje toksičnog govora ili neželjenog sadržaja) (Jha et al., 2023), za sažimanje dugih diskusija kako bi se olakšalo praćenje, ili čak za identificiranje trendova i anomalija unutar zajednice. Međutim, uspjeh ovakvih AI intervencija ovisit će o njihovoj sposobnosti da razumiju ne samo eksplicitna pravila, već i suptilne, implicitne društvene norme, kontekst i kulturne specifičnosti svake pojedine zajednice. Naučiti algoritam nijansama ljudske interakcije i društvene dinamike ostaje jedan od ključnih izazova – i jedno od najvažnijih područja istraživanja – u razvoju istinskih korisnih i odgovornih komunikacijskih agenata za podršku online zajednicama. Njihova uloga neće biti samo tehnička, već duboko društvena.

### **2.7.5 Porozne Membrane: Kad Digitalno Postane Stvarno (i Obratno)**

Nekoć se o internetu razmišljalo kao o odvojenom prostoru, kibernetičkom svijetu u koji se ulazio i iz kojeg se izlazilo. No, kako su digitalne platforme postajale sve dublje utkane u tkivo naše svakodnevice, ta se granica počela topiti, postajući sve poroznija, poput membrane kroz koju ideje, odnosi i identiteti neprestano prodiru u oba smjera. Virtualne zajednice više nisu bile samo udaljena utočišta; njihovi su se stanovnici počeli prelivati u fizički svijet, organizirajući susrete, konferencije, prosvjede i druženja koja su zamaglila razliku između online poznanika i offline prijatelja (Rainie & Wellman, 2012). Dolazak pametnih telefona, tih sveprisutnih portala u džepu koji nam osiguravaju stalnu povezanost, samo je dokinuo i posljednje iluzije o jasnoj crti razdvajanja. Naši online i offline životi više nisu paralelni svjetovi; oni su neraskidivo isprepleteni.

Platforme poput Meetup.com postale su savršen primjer ovog fenomena. One funkcioniraju kao digitalni katalizatori za fizičku interakciju, spajajući ljudе sličnih interesa – bilo da se radi o ljubiteljima rijetkih pasmina pasa, programerima koji uče novi jezik, aktivistima koji organiziraju akciju čišćenja ili članovima kluba ljubitelja knjiga. Početni kontakt i stvaranje grupe događaju se online, ali stvarni cilj je susret licem u lice: okupljanje u kafićima, parkovima, knjižnicama ili radionicama (Quan-Haase & Wellman, 2002). Ovdje se apstraktni avatari i tekstualni razgovori pretvaraju u stvarna rukovanja, zajednički smijeh i žive rasprave, demonstrirajući kako digitalno može obogatiti, a ne samo zamijeniti, fizičku društvenost i izgradnju zajednice. To je slika mrežnog individualizma gdje se veze stvaraju oko zajedničkih interesa, neovisno o geografskoj blizini, ali se često materijaliziraju u fizičkom prostoru (Wellman, 2001).

Međutim, ovo zamagljivanje granica nije prošlo bez izazivanja duboke nelagode i ozbiljnih pitanja. Dok su neki slavili nove oblike povezanosti, drugi su izražavali zabrinutost oko autentičnosti odnosa njegovanih kroz ekrane i potencijala za eroziju privatnosti u svijetu stalnog digitalnog nadzora. Je li osoba koju poznajemo kroz pažljivo kuriran online profil ista ona koju srećemo uživo? Sherry Turkle je u svojim utjecajnim radovima upozoravala na fenomen sami zajedno (alone together), gdje tehnologija stvara iluziju povezanosti dok nas zapravo izolira, potičući nas da preferiramo uredne, kontrolirane online interakcije nad neurednom nepredvidljivošću stavnih ljudskih odnosa (Turkle, 2011).

Zabrinutost oko privatnosti eksplodirala je u javnoj svijesti sa skandalima poput Cambridge Analytice, koji je razotkrio kako se osobni podaci prikupljeni s Facebooka (često bez punog razumijevanja korisnika) mogu koristiti za sofisticirano psihološko profiliranje i ciljano političko uvjerenje, potencijalno utječući na ishod izbora (Cadwalladr & Graham-Harrison, 2018). Ovaj i slični incidenti bacili su oštro svjetlo na poslovne modele mnogih platformi, utemeljene na onome što Shoshana Zuboff naziva nadzornim kapitalizmom (surveillance capitalism): ekstrakciji i monetizaciji ljudskog iskustva pretvorenenog u podatke, često kroz nejasne uvjete korištenja i sučelja dizajnirana da nas potaknu na dijeljenje više nego što bismo svjesno pristali (Zuboff, 2019). Naša online aktivnost, naši lajkovi, klikovi, pa čak i zadržavanje pogleda na određenom sadržaju, postaju sirovina za algoritme koji nas profiliraju i predviđaju naše ponašanje, oblikujući informacije koje vidimo i proizvode koji nam se nude.

Istovremeno, stalna prisutnost digitalnog svijeta u našim životima potaknula je diskusije o digitalnom otudenju, ovisnosti i utjecaju na mentalno zdravlje. Dizajn mnogih društvenih mreža, s beskonačnim skrolanjem, notifikacijama i mehanizmima varijabilne nagrade (poput lajkova), može stvoriti kompulzivne petlje ponašanja (Alter, 2017). Pritisak održavanja savršenog virtualnog identiteta, stalna usporedba s drugima i strah od propuštanja (FOMO - Fear Of Missing Out) povezani su s povećanom anksioznošću i depresijom, posebice među mlađim

korisnicima (Twenge, 2019). Granica između korištenja tehnologije kao alata i postajanja njezinim zarobljenikom postaje opasno tanka.

Dakle, dok je internet nedvojbeno srušio barijere i omogućio nove oblike zajedništva i suradnje, on je istovremeno stvorio kompleksan i često ambivalentan prostor gdje se naši digitalni i fizički životi neprestano prožimaju. Navigacija ovim hibridnim svijetom, s njegovim obećanjima i opasnostima, postala je središnji izazov modernog života, postavljajući temeljna pitanja o tome kako tehnologija oblikuje naše odnose, naš osjećaj sebe i našu percepciju stvarnosti – pitanja koja će postati još akutnija s dolaskom sve sofisticiranih komunikacijskih agenata.

### 2.7.6 Digitalna revolucija i mreža globalne komunikacije

Ulazak u digitalno doba donio je atmosferu neprekidne povezanosti i ubrzanih globalnih interakcija, pri čemu su internet i digitalni alati postali logistička i kulturna potka suvremenog društva. Zaokret od fizičke infrastrukture i tradicionalnih oblika poslovanja prema virtualnim ekosustavima stvorio je okružje u kojem se suradnja odvija neovisno o vremenu i prostoru, a raspon informacija dostupan je u opsegu koji bi do prije samo nekoliko desetljeća djelovao nezamislivo. Takva prožetost komunikacije i informacijskih tokova nastavlja dug civilizacijski niz inovacija, no sada se čini da se ritam promjena intenzivira do mjere da iz temelja preobražava mnoge aspekte ljudskog iskustva.

U kontekstu poslovanja i radnih procesa, izgradnja virtualnih tržišnih platformi ilustrira te dublje promjene. Studije o kompanijama poput Amazona i eBaya (Laudon & Traver, 2016) pokazuju kako je globalna distribucija robe i usluga ušla u novu fazu, obilježenu hibridnim modelima koji povezuju tradicionalne i digitalne modalitete trgovine. Ujedno se raširila praksa kolaboracije putem mrežnih aplikacija za razgovore, dijeljenje dokumenata i videokonferencije. Odabirom suradnika neovisno o njihovu geografskom položaju, organizacije sve više njeguju raspršene timove” u kojima se znanje stapa u jedinstvenu cjelinu (Mulki, Bardhi, Lassk, & Nanavaty-Dahl, 2009). Te se prakse paralelno prožimaju s ubrzanom kulturnom globalizacijom” u kojoj lokalna radna kultura susreće univerzalne standarde digitalne suradnje, donoseći nove izazove organizacijama koje nastoje uskladiti razne tradicije rada s fluidnošću virtualne suradnje.

Obrazovanje je doživjelo sličan zaokret. Online tečajevi i otvoreni obrazovni resursi omogućuju pristup sadržajima koji su ranije bili dostupni uglavnom u elitnim institucijama (Peters, 2003). Velika popularnost platformi poput Coursera, edX i Khan Academy (Bonk, 2009) pokazuje da je potražnja za fleksibilnim, prilagodljivim i nadasve pristupačnim oblicima učenja u stalnom porastu. Mnogi studenti i profesionalci, kojima formalni, dugotrajni studiji nisu opcija, sada biraju mikrokolegije” i specijalizirane tečajeve kako bi se brže prilagodili promjenjivim zahtjevima tržišta rada ili pratili vlastite interese u oblasti umjetnosti, znanosti ili humanistike.

Jednako su raznolike i promjene u sferi medija. Sve veća digitalna konkurenca stavila je tradicionalne novine, televizijske kuće i radio postaje pred zadatkom radikalne prilagodbe novim komunikacijskim običajima i modelima oglašavanja. S pojavom internetskih kanala, blogova, podcasta i platformi za video streaming, javio se prostor za nebrojene oblike kreativnog izražavanja. Rasprave se sada odvijaju na internetskim forumima (Rheingold, 1993), gdje se razmjenjuju stručna znanja ili nišne strasti, a informativne su emisije nerijetko zamijenjene interaktivnim prijenosima u stvarnom vremenu, čime je publika doživjela transformaciju iz pasivnog promatrača u aktivnog sudionika.

Nove ekonomski okolnosti nikako nisu izostale. Koncept ekonomije dijeljenja i platformi poput Ubera i Airbnb (Sundararajan, 2016) rasklimao je neka uvrježena pravila o vlasništvu, najmu i

radu, pokazujući koliko je digitalno posredovanje agilno u povezivanju ponude i potražnje, čak i ako to otvara pitanja regulacije i zaštite radničkih prava. Istodobno, pokreti usmjereni na otvorenu suradnju dosegnuli su vrhunac u open-source projektima, gdje globalne zajednice razvijaju složena softverska rješenja bez tradicionalnih hijerarhijskih struktura (Raymond, 1999). Takva je praksa pokazala da znanje i inovacije mogu cvjetati kada se razbijaju monopolne kontrole i kada su resursi besplatno dostupni.

Slični kolektivni mehanizmi javljaju se i u sferi financiranja. Crowdfunding kampanje putem platformi poput Kickstartera i Indiegogoa (Howe, 2008) otkrivaju kako šira publika, a ne isključivo banke ili korporativni ulagači, preuzima ulogu mikroinvestitora” u kreativnim i poduzetničkim projektima. Takve inicijative revitaliziraju kulturu osobnog angažmana, gdje podržavatelji ne samo da doniraju sredstva, već i šire vijest o novim idejama te aktivno grade zajednicu oko njih. U jednoj dimenziji to donosi osnaživanje malih” inovatora, a u drugoj nameće potrebu za vraćanjem povjerenja i transparentnosti: u digitalnom svijetu punom informacija lako je i neizvjesno procijeniti tko nudi kakvu vrijednost i s kakvim motivima.

Sve ove pojave upućuju na to da digitalna era, nastavljajući se na stoljetne tradicije komunikacijske evolucije, postaje višeslojno iskustvo — intimno i globalno, inovativno i disruptivno, obećavajuće i rizično. Omogućuje intenzivnu suradnju i velike pobjige u otkrivanju znanja, ali traži promišljene strategije upravljanja privatnošću, etikom i intelektualnim vlasništvom. Otvara nove granice za poslovanje, obrazovanje i medije, ali priziva dubinske rasprave o ravnoteži moći u digitalnom prostoru te kontroli infrastrukture i resursa. U korijenu svih tih pomaka ostaje ključno razumijevanje da je komunikacija, sada već stoljećima, nit vodilja razvoja civilizacije, a digitalne tehnologije samo su posljednji čin u neprekidnom kazalištu ljudske kreativnosti i interakcije.

#### **2.7.7 Digitalna Dvojnost: Obećanje Demokratizacije i Sjena Novih Nejednakosti**

Poput Janusa s dva lica, internet je od svog nastanka nosio **dvostruko obećanje i prijetnju**. S jedne strane, pozdravljen je kao **veliki demokratizator**, tehnologija koja će srušiti stare hijerarhije znanja i moći, pružajući svakome s pristupom mreži ključeve **globalne biblioteke, agore i izdavačke kuće u jednom** (Castells, 2001). I doista, u mnogim aspektima, internet je ispunio dio tog obećanja. Znanstvena istraživanja, nekad zaključana u skupim časopisima, postala su dostupnija kroz open-access inicijative i platforme poput arXiva. Obrazovni resursi, od Khan Academy do masovnih otvorenih online tečajeva (MOOCs), ponudili su mogućnosti učenja milijunima izvan tradicionalnih institucija. Kulturna djela, vijesti iz udaljenih kutaka svijeta i različite političke perspektive postale su dostupne na klik miša, potkopavajući kontrolu tradicionalnih čuvara vrata informacija. Građansko novinarstvo i blogosfera dali su glas onima koji su ranije bili marginalizirani u medijskom prostoru.

No, ova eksplozija informacija i povezanosti ubrzalo je otkrila i svoje **tamno naličje**. Ista infrastruktura koja omogućuje slobodan protok znanja pokazala se i kao **plodno tlo za širenje dezinformacija, propagande i manipulativnih narativa** u razmjerima i brzinom koji su prije bili nezamislivi (Marwick & Lewis, 2017; Wardle & Derakhshan, 2017). Digitalni prostor postao je zagušen informacijskim smogom, gdje je sve teže razlikovati vjerodostojne izvore od lažnih vijesti, teorija zavjere ili sofisticiranih operacija utjecaja, često pojačanih algoritamskim preporukama koje favoriziraju angažman nad točnošću. Istovremeno, personalizacija sadržaja, pokretana algoritmima koji uče naše preferencije kako bi nas zadržali na platformi (i izložili oglasima), prijeti stvaranjem **filter mješura (filter bubbles)** i **eho komora (echo chambers)** (Pariser, 2011). U tim digitalnim odajama, izloženi smo pretežno informacijama koje

potvrđuju naša postojeća uvjerenja, dok su suprotstavljena mišljenja marginalizirana, što potencijalno vodi ka **društvenoj polarizaciji i fragmentaciji zajedničke stvarnosti**. Skandali poput Cambridge Analytice dramatično su ilustrirali kako se podaci prikupljeni o našem online ponašanju mogu iskoristiti za **mikro-ciljanu manipulaciju**, bilo u komercijalne ili političke svrhe (Cadwalladr & Graham-Harrison, 2018; Zuboff, 2019).

Nadalje, obećanje univerzalnog pristupa pokazalo se iluzornim. Unatoč globalnom širenju interneta, i dalje postoje **duboke i tvrdokorne digitalne podjele (digital divides)**. Ovaj jaz nije samo pitanje fizičkog pristupa širokopojasnoj vezi ili posjedovanja adekvatnog uređaja (iako i to ostaje značajan problem u mnogim dijelovima svijeta). On uključuje i **jaz u vještina** (digitalna pismenost potrebna za kritičko vrednovanje informacija i učinkovito korištenje alata) te **jaz u korištenju** (razlika između pasivne konzumacije sadržaja i aktivnog sudjelovanja, stvaranja i iskorištavanja ekonomskih i obrazovnih prilika) (van Dijk, 2020; Hargittai, 2002). Oni koji nemaju pristup, vještine ili motivaciju za potpuno sudjelovanje u digitalnom svijetu riskiraju ostati na **margini novih ekonomskih prilika, obrazovnih resursa i demokratskih procesa**, čime se postojeće društvene i ekonomske nejednakosti ne samo reproduciraju, već i produbljuju (Robinson et al., 2015).

Unatoč ovim izazovima, internet je neosporno **transformirao dinamiku društvene i političke participacije**. Virtualne zajednice i društveni mediji postali su moćne platforme za mobilizaciju, nadilazeći geografske granice i omogućujući ljudima da se brzo okupe oko zajedničkih ciljeva, identiteta ili pritužbi. Od pokreta Zelenih u Iranu, preko Arapskog proljeća, do Occupy Wall Street, #MeToo ili Black Lives Matter, svjedočili smo kako **digitalne mreže mogu katalizirati i koordinirati kolektivne akcije** s munjevitom brzinom i globalnim dosegom (Castells, 2015; Tufekci, 2017). Hashtagovi postaju bojni poklici, a viralni sadržaj oružje u borbi za javno mnjenje. Ovo pokazuje **ogromnu emancipacijsku moć** interneta kao alata za organiziranje otpora, podizanje svijesti i traženje društvenih promjena izvan kontrole tradicionalnih institucija.

No, i ovdje se krije paradoks. Iste platforme koje omogućuju mobilizaciju mogu biti i **moćni alati za nadzor, cenzuru i kontra-mobilizaciju** od strane država i drugih moćnih aktera (Zheng et al., 2018). Aktivisti mogu biti praćeni, disidentski glasovi ušutkani, a javni diskurs manipuliran kroz koordinirane online kampanje. Zeynep Tufekci uvjerljivo argumentira da, iako digitalni alati olakšavaju brzo okupljanje velikog broja ljudi, pokreti rođeni u digitalnom dobu često nemaju organizacijsku dubinu i stratešku otpornost tradicionalnih pokreta izgrađenih kroz dugotrajni offline rad (Tufekci, 2017).

Stoga, društvene posljedice interneta ostaju duboko ambivalentne. On je istovremeno sila demokratizacije i platforma za manipulaciju; alat za premoćivanje podjela i mehanizam za stvaranje novih. On osnažuje pojedince i zajednice, ali istovremeno otvara vrata novim oblicima kontrole i nejednakosti. Razumijevanje ove složene i često kontradiktorne dinamike ključno je dok ulazimo u sljedeću fazu digitalne evolucije, gdje će umjetna inteligencija i komunikacijski agenci dodatno preoblikovati pravila igre.

#### **2.7.8 Sjene u digitalnom edenu: privatnost, manipulacija i cijena povezanosti**

Digitalni ekosustav, taj naizgled beskrajni vrt mogućnosti za širenje znanja, globalnu suradnju i ljudsko povezivanje, ima i svoju tamnu stranu, svoje skrivene ponore. Poput svakog moćnog alata, i mreža nosi inherentne rizike, a njezina sveprisutnost i kompleksnost stvaraju nove, često podmukle oblike ranjivosti. U samom temelju ovog problema leži nezasitna glad sustava za podacima. Naše online aktivnosti, od trivijalnih klikova do najintimnijih komunikacija, neprestano se bilježe, pohranjuju i analiziraju, stvarajući digitalni trag, neizbrisivi otisak našeg

postojanja koji često nismo ni svjesni da ostavljamo (Andrejevic, 2007). Tko točno upravlja tim golemlim tokovima podataka? U koje se svrhe oni koriste? Na koji način algoritamske odluke donesene na temelju naših podataka utječu na prilike koje dobivamo ili informacije koje vidimo? Odgovori su često skriveni iza neprozirnih zidova korporativnih tajni i kompleksnih uvjeta korištenja, ostavljajući korisnike u stanju informacijske asimetrije i ranjivosti (Zuboff, 2019).

Stručnjaci uporno upozoravaju: očuvanje privatnosti u digitalnom dobu nije samo pitanje osobnog komfora, već temeljno ljudsko pravo i preduvjet autonomije (Solove, 2004; Nissenbaum, 2009). Bez snažnih sigurnosnih protokola, robusne enkripcije i transparentnih mehanizama upravljanja podacima, naši digitalni identiteti postaju laka meta. Svjedočanstva o krađi identiteta, finansijskim prijevarama i neovlaštenom pristupu korisničkim računima nisu više izolirani incidenti, već dio sumorne digitalne svakodnevice. Štete mogu biti razorne, od direktnih ekonomskih gubitaka do dugotrajnog narušavanja ugleda i psihičkog stresa (Acquisti, Brandimarte & Loewenstein, 2015). Međusobna povezanost svega – od naših bankovnih računa i zdravstvenih kartona do pametnih kućanskih uređaja – stvara sve veću površinu za napad. Svaka nova poveznica u lancu potencijalna je točka probosa za sve sofisticiranije prijetnje, od ciljanih phishing napada koji nas mame na odavanje osjetljivih informacija do razornih ransomware napada koji mogu paralizirati čitave organizacije (Kshetri, 2023).

Istovremeno, otvorenost i brzina internetske komunikacije čine je idealnim vektorom za širenje dezinformacija i manipulativnih narativa. Svjedočili smo kako se ciljana politička propaganda, često zamaskirana u autentične vijesti ili građanske inicijative, koristi za utjecaj na izbore i polarizaciju javnog mnjenja (Allcott & Gentzkow, 2017; Tucker et al., 2018). Pojava deepfake tehnologije, koja omogućuje stvaranje uvjerljivih lažnih video i audio zapisa, dodaje novu, zastrašujuću dimenziju ovom problemu, priječeći potpunim urušavanjem povjerenja u ono što vidimo i čujemo (Westerlund, 2019). Društvene mreže i aplikacije za razmjenu poruka djeluju kao turbo-punjači za širenje ovakvog sadržaja; algoritmi optimizirani za angažman često favoriziraju senzacionalističke i emotivno nabijene sadržaje, bez obzira na njihovu istinitost, omogućujući lažnim vijestima da putuju brže i dalje od provjerenih informacija (Vosoughi, Roy & Aral, 2018). Borba protiv ove infodemije zahtijeva više od tehnoloških flastera poput alata za provjeru činjenica; ona traži kulturnu promjenu prema medijskoj pismenosti i kritičkom razmišljanju, sposobnosti prepoznavanja manipulativnih tehnika i vrednovanja izvora informacija kao temeljnog preduvjeta za informirano građanstvo (Vraga, Tully & Rojas, 2020).

Dodatnu razinu toksičnosti u digitalni ekosustav unosi elektroničko nasilje (cyberbullying). Za razliku od tradicionalnog nasilja, koje je često ograničeno fizičkim prostorom i vremenom, digitalno zlostavljanje može biti neumoljivo i sveprisutno. Uvredljive poruke, prijetnje, ponižavajuće slike ili glasine mogu pratiti žrtvu 24 sata dnevno, sedam dana u tjednu, prodirući u sigurnost vlastitog doma putem ekrana (Kowalski, Limber & Agatston, 2012; Hinduja & Patchin, 2015). Relativna anonimnost koju pruža digitalno okruženje može ohrabriti počinitelje, dok potencijalna trajnost i široka vidljivost štetnog sadržaja mogu stvoriti osjećaj bespomoćnosti i trajnog žiga kod žrtava. Važno je naglasiti, kako pokazuju istraživanja, da online i offline nasilje često nisu odvojeni svjetovi; počinitelji i žrtve najčešće se poznaju iz stvarnog života, iz škole ili lokalne zajednice, što ukazuje na potrebu za koordiniranim odgovorom koji uključuje roditelje, škole i širu zajednicu (Smith et al., 2008; Aboujaoude et al., 2015). Posljedice cyberbullyinga, posebice za djecu i adolescente, mogu biti razorne, uključujući pad samopoštovanja, socijalnu izolaciju, depresiju, anksioznost, pa čak i suicidalne misli (Patchin & Hinduja, 2010; John et al., 2018).

Suočeni s ovim višestrukim izazovima – erozijom privatnosti, sigurnosnim prijetnjama, poplavom dezinformacija i digitalnim nasiljem – postaje bolno jasno da tehnološki napredak sam po sebi nije jamstvo napretka. Bez adekvatnog pravnog okvira koji štiti prava pojedinaca, obrazovnih inicijativa koje promiču digitalnu pismenost i kritičko razmišljanje, te etičke svijesti ugrađene u dizajn i primjenu tehnologije, digitalni ekosustav riskira postati toksično okruženje. On je istovremeno arena nevjerojatnih mogućnosti za suradnju i inovaciju, ali i pozornica za manipulaciju, eksploraciju i sukobe. Ova digitalna revolucija, vjerojatno najbrža i najdublja komunikacijska transformacija u ljudskoj povijesti, stavlja pred nas hitan zadatak: moramo ubrzano razvijati, evaluirati i prilagodjavati pravila, norme i svijest kako bismo osigurali da tehnologija služi čovječanstvu. Samo holistički pristup, koji uravnoveže tehnološke inovacije s etičkim promišljanjem, pravnom zaštitom i obrazovanjem, može osigurati da digitalni ekosustav ostane prostor koji podržava ljudsko dostojanstvo i procvat, umjesto da postane oruđe naše vlastite dehumanizacije.

## 2.8 ZAČECI AUTONOMNE KOMUNIKACIJE: ALGORITAMSKI ŠAPAT I PRVI DIGITALNI SUGOVORNICI

Dok su prethodna poglavљa ocrtavala velike revolucije u komunikacijskim tehnologijama – od usmene riječi do globalne mreže – krajolik interneta počeo je iznjedravati nešto novo, nešto suptilnije, ali dugoročno jednako transformativno: **začetke autonomne komunikacije**. Strojevi više nisu bili samo pasivni kanali ili alati za pohranu informacija; algoritmi su počeli preuzimati aktivniju ulogu, postajući **posrednici, kuratori, pa čak i rudimentarni sudionici** u komunikacijskom procesu. Iako daleko od sofisticiranosti današnjih AI agenata, ovi rani oblici algoritamske intervencije postavili su temelje za ideju da strojevi mogu samostalno odlučivati o tome što, kada i kako komunicirati, navikavajući nas na **digitalni šapat u pozadini naših interakcija**.

Jedan od najranijih i najutjecajnijih oblika ove algoritamske medijacije pojavio se u obliku **sustava za preporučivanje (recommendation systems)**. Platforme poput Amazona, Netflix-a, Spotify-a, a kasnije i feedovi društvenih mreža, počele su koristiti algoritme kako bi nam predložile proizvode, filmove, glazbu ili sadržaj za koji su vjerovale da će nam se svidjeti (Linden, Smith, & York, 2003; Ricci, Rokach, & Shapira, 2011). Ovi sustavi, temeljeni na tehnikama poput **kolaborativnog filtriranja** (pronalazeći korisnike sličnog ukusa) ili **filtriranja temeljenog na sadržaju** (analizirajući karakteristike onoga što smo ranije konzumirali), nisu samo pasivno odgovarali na naše pretrage. Oni su **proaktivno oblikovali naš informacijski okoliš**, djelujući kao **algoritamski kuratori** koji su nam neprestano šaptali prijedloge. U određenom smislu, sustav je komunicirao svoje razumijevanje naših preferencija natrag nama, pokušavajući predvidjeti naše želje prije nego što smo ih i sami artikulirali. Iako je cilj bio komercijalni (povećanje prodaje ili zadržavanje korisnika), nuspojava je bila navikavanje na ideju da algoritam može autonomno donositi odluke o tome koji će nam se sadržaj prezentirati, sa svim kasnijim implikacijama poput filter mjeđura (Gillespie, 2014; Pariser, 2011).

Paralelno s ovim razvojem algoritamske medijacije, javili su se i prvi pokušaji stvaranja sustava koji bi **izravno simulirali ljudski razgovor: chatbotovi prve generacije**. Pionirski rad Josepha Weizenbauma sredinom 1960-ih rezultirao je programom **ELIZA**, koji je simulirao rogerijanskog psihoterapeuta (Weizenbaum, 1966). ELIZA je koristila relativno jednostavne tehnike **prepoznavanja obrazaca (pattern matching)** i transformacije korisničkog unosa kako bi generirala odgovore koji su stvarali **zapanjujuću (iako potpuno lažnu) iluziju razumijevanja i empatije**. Weizenbaum je i sam bio iznenaden, pa čak i uznemiren, koliko su ljudi bili skloni

antropomorfizirati ELIZA-u i povjeravati joj se. Iako ELIZA nije imala nikakvo stvarno razumijevanje jezika ili ljudskih emocija, ona je demonstrirala moć simulacije i posijala sjeme ideje o strojevima kao sugovornicima.

Nakon ELIZA-e, slijedili su drugi rani chatbotovi, poput **PARRY**-ja (koji je simulirao paranoidnog pacijenta) i kasnijih sustava temeljenih na jezicima poput AIML (Artificial Intelligence Markup Language), kao što je **A.L.I.C.E.** (Wallace, 2003). Ovi su sustavi često bili ograničeni na specifične domene ili su se oslanjali na goleme baze unaprijed definiranih pravila i odgovora. Njihovi razgovori bili su krhki – lako su se mogli slomiti ako bi korisnik postavio pitanje izvan očekivanog okvira ili koristio neočekivane izraze. Nedostajalo im je dublje razumijevanje konteksta, sposobnost učenja iz interakcije ili istinska fleksibilnost. Pa ipak, ovi **konverzacijiski automati** bili su važan korak. Oni su popularizirali ideju interakcije s računalom kroz prirodnji jezik i postavili temeljne izazove s kojima će se kasnije suočavati razvijatelji sofisticiranijih sustava: kako stvoriti strojeve koji mogu voditi koherentne, kontekstualno relevantne i smislene razgovore? (Dale, 2016).

Osim preporuka i ranih chatbotova, i druge algoritamske funkcije počele su djelovati kao autonomni posrednici u komunikaciji. **Spam filteri** u našim e-mail sandučićima autonomno odlučuju koje poruke zaslužuju našu pažnju, djelujući kao **digitalni vratari**. **Algoritmi za rangiranje rezultata pretrage** (poput Googleovog PageRanka i njegovih nasljednika) autonomno odlučuju koji su izvori informacija najrelevantniji za naš upit, oblikujući time naše znanje o svijetu (Pasquale, 2015). Čak i jednostavne funkcije poput **automatskog ispravljanja (autocorrect)** ili **prediktivnog teksta (predictive text)** predstavljaju oblik algoritamske intervencije u našu vlastitu komunikaciju.

Svi ovi primjeri – od algoritamskih kuratora sadržaja i ranih digitalnih sugovornika do nevidljivih filtera i rangiranja – predstavljaju **embrionalne oblike autonomne komunikacije**. Oni nisu bili agenti u današnjem smislu te riječi – nedostajala im je duboka semantička obrada, sposobnost rezoniranja, planiranja ili korištenja alata na način na koji to mogu moderni LLM agenti. Njihova autonomija često je bila ograničena na izvršavanje unaprijed definiranih pravila ili statističkih modela. Pa ipak, oni su bili ključni **prekursori**. Oni su nas polako, ali sigurno, pripremili za svijet u kojem algoritmi nisu samo pasivna infrastruktura, već **aktivni sudionici u komunikacijskoj petlji**. Pokazali su da strojevi mogu selektirati, filtrirati, pa čak i generirati komunikaciju, postavljajući pozornicu za **kvantni skok** koji će donijeti veliki jezični modeli – tehnologija koja će sposobnost strojeva da razumiju i koriste jezik podići na potpuno novu, zapanjujuću razinu, otvarajući vrata eri istinskih komunikacijskih partnera. Upravo tim skokom bavit ćemo se u sljedećem poglavljju.

### 3 SVIJET VELIKIH JEZIČNIH MODELJA: OD TEORIJE DO TEHNOLOGIJE

#### 3.1 UVOD: ULAZAK U DOBA VELIKIH JEZIČNIH MODELJA

Umetna inteligencija (AI), posebice u obliku velikih jezičnih modela (Large Language Models - LLM), predstavlja možda i najznačajniji pomak u dugo povijesti komunikacijskih tehnologija od pojave interneta. Ako promatramo razvoj ljudske komunikacije kao niz revolucija – od usmene predaje koja je gradila kolektivno pamćenje (Ong, 1982.), preko pismenosti koja je omogućila pohranu i prijenos složenog znanja (Goody & Watt, 1968.), tiskarskog stroja koji je demokratizirao pristup informacijama (Eisenstein, 1979.), elektroničkih medija koji su stvorili masovnu publiku (McLuhan, 1964.), do digitalne ere koja nas je globalno povezala (Castells, 2000.) – LLM-ovi se pojavljuju kao sila koja ne samo da posreduje u komunikaciji, već je aktivno oblikuje i generira.

Veliki jezični modeli su složeni sustavi utemeljeni na dubokim neuronским mrežama, često s milijardama ili čak trilijunima parametara, trenirani na nezamislivo velikim količinama tekstuálnih i drugih podataka (Brown et al., 2020; OpenAI, 2023). Njihova temeljna sposobnost je učenje statističkih obrazaca u jeziku do te mjere da mogu razumjeti upite i generirati koherentan, relevantan, a ponekad i iznenadujuće kreativan tekst. Jezik, taj operativni sustav ljudske misli i kulture, sada je postao izravno polje djelovanja strojeva.

Izašavši izvan okvira istraživačkih laboratorijskih radova, ova tehnologija postala je dio naše svakodnevice: pokreće chatbotove s kojima razgovaramo, prevodi jezike u stvarnom vremenu, pomaže u pisanju e-pošte, generira programski kod, sažima dugačke dokumente, pa čak i stvara umjetnička djela (Liang et al., 2022). Fascinacija njihovim mogućnostima praćena je, međutim, i dubokom zabrinutošću oko etičkih implikacija, potencijalnih pristranosti, širenja dezinformacija i utjecaja na tržište rada (Bender et al., 2021; Weidinger et al., 2021).

Ovo poglavlje zaranja u svijet velikih jezičnih modela. Istražit ćemo njihove povijesne korijene, secirati ključne tehnologije koje ih pokreću, pratiti njihov životni ciklus od sirovih podataka do praktične primjene, suočiti se s izazovima koji prate njihov razvoj i sagledati kako transformiraju krajolik ljudske komunikacije. Razumijevanje LLM-ova nije samo tehničko pitanje; ključno je za snalaženje u novom komunikacijskom dobu koje upravo nastaje.

#### 3.1 Povijesni Korijeni: Od Statistike do Dubokog Učenja

Putovanje prema današnjim sofisticiranim velikim jezičnim modelima (LLM) nije bilo linearno niti predodređeno; ono predstavlja konvergenciju desetljeća istraživanja koja se protežu kroz računalnu lingvistiku, teoriju informacija, statistiku i, naposljetku, strojno učenje s naglaskom na duboke neuronske mreže. Rani pokušaji računalnog modeliranja jezika, iako danas mogu izgledati rudimentarno, bili su ključni za postavljanje teorijskih i praktičnih temelja, identificirajući temeljne izazove u hvatanju složenosti ljudske komunikacije. Svaka nova metodologija nastojala je prevladati ograničenja prethodne, postupno gradeći slojeve razumijevanja i sposobnosti koji su kulminirali u današnjim moćnim arhitekturama.

Prvi značajni koraci u kvantitativnom modeliranju jezika oslanjali su se na statističke principe, a među najutjecajnijima bili su **n-gram modeli**. Njihova temeljna ideja, ukorijenjena u radu Claudea Shannona (1948) na teoriji informacija i Markovljevim lancima, bila je predviđanje vjerojatnosti pojavljivanja sljedeće riječi (ili znaka) u sekvenci isključivo na temelju ograničenog broja prethodnih riječi, točnije  $n-1$  riječi. Ovi modeli funkcioniраju poput kratkovidnih

prognozera koji donose odluke gledajući samo kroz vrlo mali prozor neposredne prošlosti. Unatoč svojoj konceptualnoj jednostavnosti, n-grami su se pokazali iznenadjuće korisnima za niz ranih NLP zadataka, uključujući statističko strojno prevođenje, provjeru pravopisa i rane sustave za prepoznavanje govora. Međutim, njihova fundamentalna slabost leži upravo u toj kratkovidnosti. Ljudski jezik karakteriziraju dugoročne ovisnosti – značenje riječi ili ispravnost gramatičke strukture često ovisi o kontekstu koji se nalazi mnogo ranije u tekstu. N-gram modeli, po svojoj prirodi, nisu mogli uhvatiti te udaljene veze. Štoviše, suočavali su se s problemom prokletstva dimenzionalnosti: kako se povećavao  $n$  (veličina kontekstualnog prozora), broj mogućih n-grama rastao je eksponencijalno, čineći podatke za treniranje izuzetno rijetkim (data sparsity) i modele nepraktičnima za veće kontekste. Njihovo pamćenje bilo je inherentno kratko i fragmentirano.

Potreba za prevladavanjem ograničenja fiksног i kratkог konteksta n-grama potaknula je istraživače da se okrenu **neuronskim mrežama**, posebice **rekurentnim neuronskim mrežama (RNN)**. Ideja iza RNN-ova, koju su popularizirali radovi poput onog Jeffreya Elmana (1990), bila je uvesti neku vrstu pamćenja u mrežu. Za razliku od standardnih feedforward mreža, RNN-ovi sadrže povratne veze (cikluse), omogućujući informacijama da opstaju kroz korake obrade sekvene. To se postiže kroz koncept **skrivenog stanja (hidden state)**, vektora koji u svakom koraku sažima relevantne informacije iz prethodnih dijelova sekvene i prenosi ih u sljedeći korak. Teoretski, ovo je RNN-ovima omogućilo obradu sekvenci varijabilne duljine i hvatanje ovisnosti između elemenata neovisno o njihovoј udaljenosti. Mogli bismo ih zamisliti kao čitatelje koji kontinuirano ažuriraju svoje mentalno stanje dok prolaze kroz tekst, pokušavajući zapamtiti ključne informacije. Međutim, ova elegantna teorijska ideja suočila se s ozbiljnim praktičnim preprekama tijekom procesa treniranja, poznatim kao problem **nestajućih (vanishing) i eksplodirajućih (exploding) gradijenata** (Bengio et al., 1994). Prilikom propagiranja gradijenata pogreške unatrag kroz vrijeme (backpropagation through time), gradijenti su ili eksponencijalno opadali prema nuli (nestajanje) ili eksponencijalno rasli (eksplodiranje). Nestajući gradijenti učinili su gotovo nemogućim da mreža nauči dugoročne ovisnosti, jer signal pogreške s kraja duge sekvene nije mogao efikasno doprijeti do početnih koraka kako bi prilagodio relevantne težine. Informacija se, stikovito rečeno, gubila ili iskrivljivala na dugom putu kroz rekurentne veze.

Značajan probaj u rješavanju problema nestajućih gradijenata i omogućavanju učenja dugoročnih ovisnosti došao je s razvojem sofisticiranih rekurentnih arhitektura: **mreža s Dugim Kratkoročnim Pamćenjem (Long Short-Term Memory - LSTM)**, koje su predstavili Sepp Hochreiter i Jürgen Schmidhuber (1997), te kasnije nešto jednostavnijih **Jedinica s Vratima za Rekurenkciju (Gated Recurrent Units - GRU)** (Cho et al., 2014). Ključna inovacija ovih arhitektura bila je uvođenje **mehanizama vrata (gates)** – specijaliziranih neuronskih komponenti koje dinamički kontroliraju protok informacija unutar rekurentne jedinice. LSTM jedinica, primjerice, koristi tri glavna vrata: **ulazna vrata (input gate)** koja odlučuju koje nove informacije treba pohraniti u stanje ćelije, **zaboravna vrata (forget gate)** koja odlučuju koje stare informacije treba odbaciti iz stanja ćelije, i **izlazna vrata (output gate)** koja odlučuju koji dio stanja ćelije treba koristiti za izračun skrivenog stanja (izlaza jedinice). GRU pojednostavljuje ovu arhitekturu koristeći samo dvoja vrata (update i reset). Ova sposobnost selektivnog pamćenja relevantnih informacija i zaboravljanja nebitnih omogućila je LSTM-ovima i GRU-ovima da uspešno nauče ovisnosti koje se protežu kroz znatno duže sekvene nego što je to bilo moguće s jednostavnim RNN-ovima. Metaforički, to je kao da je naš čitatelj dobio sofisticirani sustav za podcrtavanje, bilježenje i brisanje informacija, omogućujući mu da zadrži ključne elemente radnje ili argumenta kroz cijelu knjigu. Zahvaljujući ovim sposobnostima,

LSTM i GRU mreže dominirale su područjem obrade prirodnog jezika (NLP) dugi niz godina, postižući vrhunske rezultate u zadacima poput strojnog prevodenja, analize sentimenta, generiranja teksta i prepoznavanja govora.

Ipak, čak su i LSTM i GRU imali svoja ograničenja. Njihova inherentno sekvencijalna priroda obrade – riječ po riječ – otežavala je **paralelizaciju** procesa treniranja na modernom hardveru (poput GPU-ova) i još uvek je predstavljala izazov za hvatanje vrlo dugih ovisnosti u ekstremno dugim dokumentima. Prava **revolucija** dogodila se 2017. godine s objavom rada naslovjenog **Attention Is All You Need** (Vaswani et al., 2017). Ovaj rad predstavio je **Transformer arhitekturu**, koja je radikalno odstupila od rekurentnog pristupa. Umjesto sekvencijalne obrade, Transformeri se u potpunosti oslanjaju na **mehanizam pažnje (attention)**, posebice **samopažnju (self-attention)**. Samopažnja omogućuje modelu da, prilikom obrade jedne riječi (ili tokena) u ulaznoj sekvenci, direktno pogleda i izračuna relevantnost *svih ostalih* riječi u istoj sekvenci, neovisno o njihovoj udaljenosti. Za svaku riječ, model uči generirati tri vektora: Upit (Query), Ključ (Key) i Vrijednost (Value). Izračunavanjem skalarnog produkta između Upita trenutne riječi i Ključeva svih ostalih riječi, te primjenom softmax funkcije, dobivaju se skorovi pažnje koji predstavljaju težine. Te težine se zatim koriste za izračun ponderiranog zbroja Vrijednosti svih riječi, dajući kontekstualiziranu reprezentaciju trenutne riječi koja uzima u obzir cijelu sekvencu. To je kao da čitatelj može jednim pogledom obuhvatiti cijelu stranicu ili poglavljje i trenutno uspostaviti veze između svih relevantnih pojmoveva, bez potrebe da sekvencijalno prelazi tekst. Sposobnost direktnog modeliranja odnosa između svih parova riječi pokazala se izuzetno moćnom za hvatanje dugoročnih ovisnosti. Jednako važno, izračuni unutar Transformer slojeva mogu se u velikoj mjeri paralelizirati, što je omogućilo treniranje znatno većih modela na masivnim skupovima podataka u razumnom vremenu. Superiorne performanse u hvatanju konteksta i mogućnost skaliranja učinile su Transformere de facto standardom i temeljem gotovo svih modernih velikih jezičnih modела.

Na čvrstim temeljima Transformer arhitekture ubrzo su izgrađeni modeli koji su redefinirali stanje tehnike u NLP-u i privukli ogromnu pažnju znanstvene zajednice i javnosti. Dva najutjecajnija pravca predstavljaju **BERT** i **GPT** serija modela.

**BERT (Bidirectional Encoder Representations from Transformers)**, predstavljen od strane Googlea (Devlin et al., 2019), bio je revolucionaran zbog svoje **dvosmjerne** prirode pre-treniranja. Za razliku od ranijih popularnih modela (uključujući rane GPT modele) koji su bili autoregresivni i gledali samo lijevi (prethodni) kontekst prilikom predviđanja sljedeće riječi, BERT je koristio arhitekturu temeljenu isključivo na Transformer enkoderima, a bio je predtreniran na dva zadatka: **Maskirano modeliranje jezika (Masked Language Model - MLM)**, gdje je model morao predvidjeti nasumično maskirane (sakrivene) riječi unutar rečenice koristeći *i lijevi i desni* kontekst, te zadatak **Predviđanje sljedeće rečenice (Next Sentence Prediction - NSP)**, gdje je model učio razumjeti odnos između parova rečenica. Dvosmjerni kontekst omogućen MLM zadatkom dao je BERT-u znatno dublje razumijevanje značenja riječi unutar nijihovog punog konteksta. BERT je postigao vrhunske rezultate na širokom spektru NLU (Natural Language Understanding) zadataka, kao što su odgovaranje na pitanja, analiza sentimenta i prepoznavanje imenovanih entiteta, postavši standard za evaluaciju na benchmarkovima poput GLUE (Wang et al., 2018).

S druge strane, **GPT (Generative Pretrained Transformer)** serija modela, razvijena od strane tvrtke OpenAI, fokusirala se primarno na **generativne sposobnosti (NLG - Natural Language Generation)**. Koristeći arhitekturu temeljenu isključivo na Transformer dekoderima, GPT modeli su pre-trenirani na klasičnom zadatku modeliranja jezika – predviđanju sljedeće riječi u

sekvenci, što ih čini inherentno **autoregresivnim** ili jednosmjernim. Njihova evolucija demonstrira nevjerljivo učinkne skaliranja: **GPT-1** (Radford et al., 2018) pokazao je potencijal Transformer dekodera za generativne zadatke; **GPT-2** (Radford et al., 2019) iznenadio je zajednicu svojom sposobnošću generiranja iznenađujuće koherenčnih i stilski raznolikih tekstova na temelju kratkog prompta, pokazujući snažne **zero-shot** sposobnosti (rješavanje zadataka bez specifičnih primjera); **GPT-3** (Brown et al., 2020), sa svojih 175 milijardi parametara, podigao je ljestvicu demonstrirajući zapanjujuće **few-shot** (rješavanje zadataka s vrlo malo primjera) i zero-shot performanse na širokom rasponu zadataka, te pokazujući **emergentne sposobnosti** – nove sposobnosti koje se pojavljuju tek na određenoj skali modela. Konačno, **GPT-4** (OpenAI, 2023) i njegov nasljednik **GPT-4o** (OpenAI, 2024b) dodatno su poboljšali performanse u razumijevanju, rezoniranju, kodiranju i, što je ključno, uveli snažne **multimodalne** sposobnosti, omogućujući obradu i generiranje ne samo teksta, već i slika i zvuka u stvarnom vremenu. Upravo je GPT serija, posebice kroz sučelja poput ChatGPT-a, popularizirala velike jezične modele i dovela ih iz akademskih krugova u mainstream svijest.

Ovaj povijesni pregled, od jednostavnih statističkih n-grama do masivnih, Transformer-baziranih modela poput BERT-a i GPT-a, ilustrira dinamičnu evoluciju polja. Svaki korak predstavlja je odgovor na ograničenja prethodnih pristupa, uvodeći nove koncepte i arhitekture koje su omogućile sve dublje razumijevanje i generiranje ljudskog jezika. Razumijevanje ovog puta ključno je ne samo za cijenjenje trenutnih sposobnosti LLM-ova, već i za prepoznavanje njihovih inherentnih ograničenja i usmjeravanje budućih istraživanja prema još naprednjim i pouzdanim sustavima umjetne inteligencije.

### 3.2 ANATOMIJA LLM-A: KLUČNE TEHNOLOGIJE I ARHITEKTURE

Da bismo istinski razumjeli izvanredne sposobnosti velikih jezičnih modela u obradi i generiranju ljudskog jezika, nužno je zaviriti dublje u njihovu unutarnju strukturu i temeljne tehnološke mehanizme. Iako su potpuni detalji njihove implementacije iznimno složeni i predstavljaju vrhunac suvremenog inženjerstva umjetne inteligencije, dvije temeljne komponente čine okosnicu gotovo svih modernih LLM-ova: sofisticirane metode za **vektorsku reprezentaciju riječi (embeddings)** koje prevode jezik u numerički oblik razumljiv strojevima, te revolucionarna **arhitektura Transformer-a** koja omogućuje modeliranje složenih odnosa unutar jezičnih sekvenci. Upravo sinergija ovih dvaju elemenata omogućuje LLM-ovima da dosegnu razinu performansi koja je donedavno bila nezamisliva.

Prvi fundamentalni izazov u računalnoj obradi jezika jest premošćivanje jaza između simboličke prirode ljudskog jezika i numeričke prirode računalnih operacija. Računala ne operiraju riječima, već brojevima. Stoga je ključni prvi korak pretvaranje riječi ili manjih jezičnih jedinica (tokena) u numeričke reprezentacije. Rani pristupi koristili su jednostavne metode poput one-hot kodiranja, gdje se svaka riječ predstavlja dugačkim vektorom ispunjenim nulama, s jedinicom na poziciji koja odgovara toj riječi u rječniku. Međutim, ovakve reprezentacije su izuzetno rijetke (sparse), visokodimenzionalne i, što je najvažnije, ne hvataju nikakvu semantičku sličnost između riječi – vektori za pas i mačku bili bi jednakо udaljeni kao vektori za pas i stolica.

Značajan napredak postignut je uvođenjem **gustih vektorskih reprezentacija (dense embeddings)**, poznatih i kao **uložišni vektori (embeddings)**. Ideja je predstaviti svaku riječ kao relativno kratak vektor realnih brojeva (npr. dimenzija 100 do nekoliko tisuća) u kontinuiranom višedimenzionalnom prostoru, pri čemu se vektori uče tako da odražavaju semantička i sintaktička svojstva riječi iz velikih korpusa teksta. Cilj je da riječi sa sličnim značenjem ili one

koje se pojavljuju u sličnim kontekstima imaju vektore koji su geometrijski blizu jedan drugome u tom vektorskom prostoru. Pionirski radovi poput **Word2Vec** (Mikolov et al., 2013), koji je uveo dvije efikasne arhitekture – Continuous Bag-of-Words (CBOW, predviđanje riječi iz okolnog konteksta) i Skip-gram (predviđanje okolnog konteksta iz riječi) – te **GloVe (Global Vectors for Word Representation)** (Pennington et al., 2014), koji je kombinirao prednosti lokalnog kontekstualnog prozora (kao u Word2Vec) s globalnom statistikom supojavljivanja riječi u korpusu, revolucionirali su područje. Ovi modeli uspješno su hvatali semantičke sličnosti (npr. blizina vektora za pas i mačka) pa čak i analogijske odnose (npr. vektor(kralj) - vektor(muškarac) + vektor(žena)  $\approx$  vektor(kraljica)). Međutim, ovi rani pristupi generirali su **statičke embeddingse**: svaka riječ u rječniku imala je točno jedan, fiksni vektorski prikaz, neovisno o kontekstu u kojem se pojavljuje. Ovo je predstavljalo značajno ograničenje jer ljudski jezik obiluje **polisemijom** – ista riječ može imati različita značenja ovisno o kontekstu (npr. engleska riječ bank može označavati finansijsku instituciju ili obalu rijeke). Statički embeddingsi nisu mogli razriješiti ovu dvosmistenost.

Potreba za modeliranjem značenja riječi u specifičnom kontekstu dovela je do razvoja **kontekstualiziranih embeddinga**. Umjesto da se svakoj riječi dodijeli fiksni vektor, kontekstualizirani pristupi generiraju reprezentaciju za svaku pojavu (token) riječi dinamički, uzimajući u obzir okolne riječi u rečenici ili dokumentu. Rani primjeri uključuju ELMo (Embeddings from Language Models) (Peters et al., 2018), koji je koristio dvostrane LSTM mreže za generiranje embeddinga koji su funkcija cijele ulazne rečenice. Međutim, pravi procvat kontekstualiziranih embeddinga dogodio se s pojavom Transformer arhitekture. Modeli poput BERT-a (Devlin et al., 2019) i GPT serije (Radford et al., 2018, 2019; Brown et al., 2020) ne koriste predefinirane embeddinge kao ulaz u fiksnom smislu; umjesto toga, sami duboki slojevi Transformera generiraju bogate, kontekstualizirane reprezentacije za svaki token kao dio procesa obrade. "Vektorski prikaz riječi **kosa** u rečenici 'Njegovala je svoju dugu, sjajnu **kosu**' (misleći na vlasni glavi) bit će značajno drugačiji od onoga za istu riječ u rečenici 'Poljoprivrednik je naoštirovao **kosu** prije izlaska na livadu' (misleći na alat za košenje)." Ovi dinamički, kontekstualno ovisni vektori, koji proizlaze iz interakcija unutar složenih neuronskih slojeva, omogućuju LLM-ovima mnogo nijansiranije i preciznije razumijevanje semantike jezika, uspješno rješavajući problem polisemije i hvatajući suptilne razlike u značenju.

Druga ključna komponenta, i vjerojatno najvažnija inovacija koja pokreće moderne LLM-ove, jest **arhitektura Transformer-a**, predstavljena u seminarном radu "Attention Is All You Need" (Vaswani et al., 2017). Ova arhitektura napušta rekurentne veze koje su dominirale sekvencijskim modeliranjem (poput RNN-ova i LSTM-ova) i umjesto toga se u potpunosti oslanja na **mehanizam pažnje (attention mechanism)** za modeliranje ovisnosti između različitih dijelova ulazne i izlazne sekvence. Standardna Transformer arhitektura sastoji se od dva glavna dijела: **enkoderskog (encoder)** i **dekoderskog (decoder)** sklopa (stack).

**Enkoder** ima zadatak primiti ulaznu sekvencu (npr. rečenicu na izvornom jeziku u zadatu prevođenja) i transformirati je u niz kontinuiranih, kontekstualiziranih reprezentacija, po jednu za svaki ulazni token. Svaki enkoder u sklopu (obično ih je više naslaganih jedan na drugi) sastoji se od dva glavna podsloja: **višeglave samoponažnje (multi-head self-attention)** i **potpuno povezane prednaponske mreže (position-wise fully connected feed-forward network)**. Oko svakog od ova dva podsloja primjenjuju se **rezidualne veze (residual connections)** (He et al., 2016), koje pomažu u ublažavanju problema nestajućih gradijenata u vrlo dubokim mrežama, nakon čega slijedi **normalizacija sloja (layer normalization)** (Ba et al., 2016), koja stabilizira proces treniranja. Sloj samoponažnje omogućuje enkoderu da za svaki

token u ulaznoj sekvenci odredi koliko pažnje treba posvetiti svim ostalim tokenima u istoj sekvenci prilikom izračuna njegove reprezentacije, efektivno hvatajući unutarnje ovisnosti i kontekst unutar ulaza. Prednaponska mreža zatim dodatno obrađuje svaku poziciju neovisno.

**Dekoder** ima zadatak generirati izlaznu sekvencu (npr. prijevod na ciljni jezik), token po token. Slično enkoderu, dekoder se također sastoji od više naslaganih identičnih slojeva. Svaki dekoderski sloj, uz dva podsloja prisutna u enkoderu (samopažnja i prednaponska mreža), sadrži i treći ključni podsloj: **višeglavu unakrsnu pažnju (multi-head cross-attention)**. Samopažnja u dekoderu je **maskirana (masked self-attention)** – prilikom predviđanja trenutnog tokena, modelu je dopušteno obraćati pažnju samo na prethodne tokene u generiranoj sekvenci i na sam trenutni token, kako bi se sprječilo "gledanje u budućnost" i očuvala autoregresivna priroda generiranja. Ključni novi element je unakrsna pažnja, koja omogućuje dekoderu da obrati pažnju na izlazne reprezentacije generirane od strane enkodera. To jest, dok generira svaki izlazni token, dekoder može pogledati cijelu kodiranu reprezentaciju ulazne sekvence i fokusirati se na one dijelove ulaza koji su najrelevantniji za generiranje trenutnog izlaznog tokena. Kao i u enkoderu, koriste se rezidualne veze i normalizacija sloja.

Centralni mehanizam koji omogućuje svu ovu funkcionalnost jest **pažnja (attention)**, posebice **skalirana pažnja s točkastim produkтом (Scaled Dot-Product Attention)**, koja čini osnovu samopažnje i unakrsne pažnje. Za svaki element u sekvenci (npr. token), model uči tri vektora: **Upit (Query - Q)**, **Ključ (Key - K)** i **Vrijednost (Value - V)**. Intuitivno, Upit predstavlja trenutni fokus ili pitanje koje postavlja element; Ključ predstavlja neku vrstu "oznake" ili "indeksa" za informaciju koju nosi svaki element; Vrijednost predstavlja samu informaciju koju nosi element. Mehanizam pažnje izračunava sličnost između Upita trenutnog elementa i Ključeva svih ostalih elemenata (obično koristeći skalarni produkt). Ovi skorovi sličnosti se zatim skaliraju (kako bi se sprječile vrlo velike vrijednosti koje bi mogle destabilizirati softmax funkciju) i propuštaju kroz softmax funkciju kako bi se dobile **težine pažnje (attention weights)** – distribucija vjerojatnosti koja pokazuje koliko je svaki element relevantan za trenutni Upit. Konačno, izlaz pažnje za trenutni element dobiva se kao ponderirani zbroj Vrijednosti svih elemenata, gdje su težine upravo izračunate težine pažnje. Na taj način, reprezentacija svakog elementa postaje obogaćena informacijama iz drugih, relevantnih dijelova sekvence.

Transformer arhitektura dodatno poboljšava ovaj mehanizam koristeći **višeglavu pažnju (Multi-Head Attention)**. Umjesto da izračunava pažnju samo jednom s punim dimenzijama Q, K i V vektora, model prvo linearno projicira Q, K i V vektore  $h$  puta (gdje je  $h$  broj "glava") u prostoru nižih dimenzija. Zatim se mehanizam skalirane pažnje s točkastim produkтом primjenjuje paralelno za svaku od ovih projiciranih verzija ("glava"). Svaka glava može naučiti fokusirati se na različite aspekte odnosa ili različite reprezentacijske podprostore. Izlazi svih glava se zatim konkatentiraju (spajaju) i ponovno linearne projiciraju kako bi se dobio konačni izlaz višeglave pažnje. Ovo omogućuje modelu da "zajednički obrati pažnju na informacije iz različitih reprezentacijskih podprostora na različitim pozicijama" (Vaswani et al., 2017, str. 5), što rezultira bogatijom i snažnjom reprezentacijom.

**Samopažnja (self-attention)** je slučaj kada Q, K i V vektori dolaze iz iste sekvence (npr. unutar enkodera ili unutar dekodera, gdje model povezuje različite dijelove iste rečenice). Metafora povezivanja zamjenice "ona" s odgovarajućom imenicom ranije u tekstu ilustrira moć samopažnje u hvatanju unutarnjih referenci i ovisnosti. **Unakrsna pažnja (cross-attention)** javlja se u dekoderu kada Q vektori dolaze iz dekoderske sekvence (generirani tekst), dok K i V vektori dolaze iz enkoderske sekvence (ulazni tekst). Ovo omogućuje dekoderu da, dok generira prijevod ili odgovor, stalno konzultira relevantne dijelove izvornog ulaza.

Upravo ova moćna kombinacija Transformer arhitekture, koja napušta sekvencijalna uska grla rekurentnih modela u korist visoko paralelizabilnog mehanizma pažnje sposobnog za hvatanje dugoročnih ovisnosti, i sofisticiranih, dinamički generiranih kontekstualiziranih embeddinga, daje velikim jezičnim modelima njihovu sposobnost dubokog semantičkog razumijevanja i generiranja fluentnog, koherentnog i kontekstualno relevantnog jezika na razini koja neprestano pomiče granice umjetne inteligencije.

### 3.3 ŽIVOTNI CIKLUS MODELA: OD PODATAKA DO PRIMJENE

Stvaranje i korištenje velikog jezičnog modela složen je proces koji se može podijeliti u nekoliko ključnih faza:

#### 3.3.1 Pre-treniranje (Pre-Training): Stvaranje Temeljnog Znanja

Faza pre-treniranja predstavlja temelj na kojem se grade sposobnosti modernih velikih jezičnih modela i ujedno je najzahtjevniji korak u smislu potrebnih računalnih resursa, ogromnih količina podataka i uloženog vremena. U ovoj inicijalnoj, ali ključnoj fazi, model se izlaže masivnim, uglavnom nestrukturiranim korpusima tekstualnih podataka – često reda veličine terabajta ili čak petabajta – prikupljenih iz raznolikih izvora kao što su prostranstva weba (npr. koristeći podatke iz Common Crawl-a), digitalizirane knjižnice (knjige), znanstveni članci (npr. arXiv), repozitoriji koda (npr. GitHub) i drugi dostupni digitalni tekstovi (Brown et al., 2020; Touvron et al., 2023; Rae et al., 2021). Sve više, ovi korupsi uključuju i podatke drugih modaliteta, poput slika i zvuka, kako bi se razvili multimodalni modeli. Primarni cilj pre-treniranja nije osposobiti model za izvršavanje nekog specifičnog, unaprijed definiranog zadatka (kao što je prevodenje ili odgovaranje na pitanja), već mu omogućiti da, kroz proces učenja na ovim golemlim podacima, samostalno otkrije i internalizira temeljne statističke obrasce, strukture i zakonitosti ljudskog jezika. Ovo uključuje učenje gramatike, sintakse, semantičkih odnosa između riječi i koncepata, stilističkih nijansi, pa čak i stjecanje značajne količine činjeničnog znanja o svijetu (Petroni et al., 2019), isključivo kroz analizu načina na koji se jezik koristi u promatranim podacima.

Ključna paradigma koja omogućuje učenje iz ovako velikih, ali uglavnom neoznačenih (unlabeled) skupova podataka jest **samonadzirano učenje (self-supervised learning)**. Za razliku od klasičnog nadziranog učenja, gdje model uči mapirati ulaze na izlaze na temelju eksplicitno pruženih oznaka (npr. parova rečenica i njihovih prijevoda, ili slika i njihovih kategorija), samonadzirano učenje koristi inherentnu strukturu samih podataka za automatsko generiranje ciljeva učenja. Modelu se ne daju vanjske "točne" oznake; umjesto toga, dio ulaznih podataka se namjerno sakriva ili modificira, a zadatak modela je rekonstruirati ili predvidjeti taj skriveni dio na temelju ostatka vidljivih podataka. Ova sposobnost generiranja vlastitih nadzornih signala iz sirovih podataka ključna je za iskoristavanje ogromnih količina teksta dostupnih na internetu, za koje bi ručno označavanje bilo potpuno nepraktično i preskupo.

U kontekstu pre-treniranja velikih jezičnih modela, dominiraju dva glavna samonadzirana zadataka (objectives):

1. **Modeliranje jezika (Language Modeling - LM):** Ovo je klasični zadatak u obradi prirodnog jezika, a u kontekstu autoregresivnih modela poput GPT serije (Radford et al., 2018, 2019; Brown et al., 2020), često se naziva **kauzalno modeliranje jezika (Causal Language Modeling - CLM)**. Zadatak modela je predvidjeti sljedeći token (rijec ili dio riječi) u sekvenci, s obzirom na sve prethodne tokene. Matematički, model uči distribuciju vjerojatnosti  $P(\text{token}_i | \text{token}_1, \dots, \text{token}_{i-1})$ . Iako se zadatak čini

jednostavnim (predviđanje sljedeće riječi), da bi ga uspješno obavljao na velikom i raznolikom korpusu, model mora implicitno naučiti izuzetno složene aspekte jezika: gramatička pravila koja određuju koje su sekvence riječi vjerojatne, semantičke odnose koji povezuju značenja riječi, kontekstualne ovisnosti koje se protežu kroz rečenice i odlomke, pa čak i činjenično znanje potrebno za predviđanje smislenih nastavaka (npr. predvidjeti "Pariz" nakon "Glavni grad Francuske je..."). Ovaj pristup je inherentno generativan i čini temelj modela koji su izvrsni u stvaranju teksta.

2. **Maskirano modeliranje jezika (Masked Language Modeling - MLM):** Ovaj zadatak, populariziran od strane BERT modela (Devlin et al., 2019), koristi drugačiji pristup. Umjesto predviđanja sljedećeg tokena, nasumični postotak tokena u ulaznoj sekvenci se zamjenjuje posebnim "[MASK]" tokenom (ili se ponekad zamjenjuje drugim nasumičnim tokenom ili ostavlja nepromijenjen kako bi se smanjio raskorak između pre-treniranja i finog podešavanja). Zadatak modela je zatim predvidjeti originalne identitete tih maskiranih tokena, ali za razliku od CLM-a, pri tome može koristiti **cjelokupni kontekst** – i tokene koji dolaze prije i tokene koji dolaze poslije maskiranog tokena (dvosmerni kontekst). Ova sposobnost uvjetovanja na oba smjera omogućuje modelu da razvije dublje i nijansirane razumijevanje kontekstualnog značenja riječi, što se pokazalo posebno korisnim za zadatke razumijevanja prirodnog jezika (NLU). Varijacije ovog pristupa uključuju i druge "denoising" (uklanjanje šuma) ciljeve, poput predviđanja izbrisanih dijelova teksta (spans) kao u T5 modelu (Raffel et al., 2020).

Izbor pre-trenirajućeg zadatka (ili njihove kombinacije) značajno utječe na arhitekturu i sposobnosti rezultirajućeg modela (npr. GPT modeli koriste samo dekoderske blokove Transformera zbog autoregresivne prirode CLM-a, dok BERT koristi samo enkoderske blokove za MLM).

Jedan od najvažnijih empirijskih uvida koji je pokrenuo eksploziju velikih jezičnih modela jesu **zakoni skaliranja (scaling laws)**. Istraživanja, posebice ona iz OpenAI-ja (Kaplan et al., 2020) i DeepMind-a (Hoffmann et al., 2022), pokazala su da performanse (mjerene kao gubitak na zadatku modeliranja jezika) velikih jezičnih modela predvidljivo i glatko poboljšavaju kao funkcija tri ključna faktora: **veličine modela** (broja parametara), **veličine skupa podataka** za pre-treniranje, i **količine računalnih resursa** (compute) utrošenih na treniranje. Ovi zakoni sugeriraju da, unutar određenih granica, jednostavno povećanje ovih faktora dovodi do boljih modela. Posebno je značajan rad Hoffmanna i suradnika (2022) koji sugerira da su mnogi prethodni veliki modeli bili "nedovoljno trenirani" – bili su preveliki za količinu podataka na kojoj su trenirani. Predložili su optimalne omjere između veličine modela i podataka za dani računalni budžet (tzv. "Chinchilla scaling laws"), implicirajući da je za postizanje najboljih performansi potrebno proporcionalno povećavati i veličinu modela i količinu podataka. Ovo otkriće potaknulo je utruku prema sve većim modelima i prikupljanju još masivnijih skupova podataka. Fascinantna posljedica skaliranja je pojava **emergentnih sposobnosti** (Wei et al., 2022b) – sposobnosti (npr. rješavanje aritmetičkih zadataka, odgovaranje na pitanja u određenim formatima, few-shot učenje) koje nisu prisutne ili su vrlo slabe u manjim modelima, ali se relativno naglo pojavljuju kada veličina modela prijeđe određeni prag. Prijeklo ovih sposobnosti još uvjek nije potpuno razjašnjeno, ali one dodatno motiviraju istraživanje na ekstremnim skalamama.

Uspjeh pre-treniranja, međutim, ne ovisi isključivo o kvantiteti, već presudno i o **kvaliteti, raznolikosti i etičkom promišljanju izvora podataka**. Iako se koriste masivni skupovi poput

Common Crawl-a, koji obuhvaćaju petabajte sirovog teksta s interneta, ti su podaci inherentno heterogeni i "bučni". Oni neizbjješno sadrže značajne količine niskokvalitetnog, repetitivnog, algoritamski generiranog, pa čak i toksičnog ili eksplicitnog sadržaja. Što je još važnije za kasnije komunikacijske primjene, ovi korpsi su **zrcalo ljudskog društva online**, sa svim njegovim ukorijenjenim **društvenim pristranostima** – rasnim, rodnim, kulturnim, socioekonomskim i drugim stereotipima koji se manifestiraju u jeziku (Blodgett et al., 2020). Nadalje, agregacija podataka s weba povlači i kompleksna pitanja vezana uz **privatnost**, jer može sadržavati osjetljive osobne informacije, te pitanja **autorskih prava** (Liang et al., 2021).

Stoga je faza **rigorozne obrade i filtriranja podataka** prije samog pre-treniranja od apsolutne nužnosti, predstavljajući ključni korak u oblikovanju temeljnih sposobnosti i, što je još važnije, potencijalnih nedostataka budućeg modela (Rae et al., 2021; Touvron et al., 2023). Ovaj višefazni proces tipično uključuje: primjenu heuristika ili modela za klasifikaciju kvalitete kako bi se **filtrirao niskokvalitetni sadržaj**; korištenje specijaliziranih klasifikatora ili opsežnih lista ključnih riječi za **identifikaciju i uklanjanje štetnog ili toksičnog materijala**; agresivnu **deduplikaciju** na različitim razinama granularnosti (od cijelih dokumenata do pojedinačnih rečenica ili n-grama) kako bi se sprječilo prekomjerno učenje repetitivnih sekvenci i smanjila nemamjerna memorizacija specifičnih podataka (Lee et al., 2021); te napore u **detekciji i uklanjanju ili maskiranju osobnih identifikacijskih informacija (PII)**, iako je postizanje potpune anonimizacije izuzetno izazovno (Carlini et al., 2021). Konačno, pažljivo **balansiranje i miješanje (Data Mixing)** podataka iz različitih izvora (npr. web tekst, knjige, znanstveni radovi, kod) ključno je za postizanje želenog profila znanja i sposobnosti u rezultirajućem modelu.

Unatoč svim naporima u kuriranju i čišćenju podataka, fundamentalno je važno naglasiti da **pristranosti prisutne u filtriranim, ali i dalje inherentno pristranim izvornim podacima, neizbjježno bivaju internalizirane i često čak i pojačane od strane modela tijekom pre-treniranja** (Bender et al., 2021; Sheng et al., 2021). Model uči jezične obrasce, ali zajedno s njima uči i asocijacije i stereotipe ugrađene u te obrasce. Ovo predstavlja jedan od najznačajnijih i najtvrdokornijih etičkih izazova u razvoju LLM-ova, jer direktno utječe na potencijal modela da kasnije, u ulozi komunikacijskog agenta, generira **pristran, nepravedan ili stereotipan jezik**. Borba protiv ovih naučenih pristranosti zahtijeva kontinuirane napore ne samo u filtriranju podataka, već i u kasnijim fazama finog podešavanja i poravnjanja.

Konačni ishod faze pre-treniranja je **temeljni model (foundation model)**, termin populariziran od strane Stanfordovog instituta za AI usmjerjen na čovjeka (Bommasani et al., 2021). Ovaj model posjeduje široko, opće razumijevanje jezika i implicitno znanje o svijetu, destilirano iz ogromnih količina procesiranih podataka. On predstavlja moćnu platformu koja se, kroz daljnje faze životnog ciklusa, može prilagoditi (tipično finim podešavanjem) za izvršavanje impresivnog niza različitih nizvodnih zadataka, uključujući one ključne za komunikacijske agente poput odgovaranja na pitanja, sažimanja ili prevodenja. Međutim, važno je razumjeti da temeljni model, odmah nakon pre-treniranja, još uvijek nije nužno optimiziran za sigurnu ili kooperativnu interakciju s ljudima; on je moćna, ali još uvijek **"sirova"** i **neusklađena osnova**. Faza pre-treniranja tako postavlja temelje i definira inherentne sposobnosti i ograničenja, ali put do pouzdanog i korisnog komunikacijskog partnera zahtijeva daljnju obradu i pažljivo usmjeravanje.

### 3.3.2 Fino Podešavanje i Poravnanje (Fine-Tuning & Alignment): Prilagodba Svrsi

Nakon intenzivne faze pre-treniranja, rezultirajući temeljni model (foundation model) posjeduje impresivnu širinu lingvističkog znanja i određenu količinu enciklopedijskog znanja o svijetu, naučenu iz ogromnih korpusa podataka (Bommasani et al., 2021). Međutim, ovaj model je poput izuzetno obrazovanog, ali nespecijaliziranog i potencijalno nepredvidljivog entiteta. Njegove sirove sposobnosti, iako općenite, nisu nužno optimizirane za specifične zadatke koje korisnici žele obavljati, niti je njegovo ponašanje nužno uskladeno s ljudskim očekivanjima o korisnosti, istinitosti i sigurnosti. Stoga, da bi se premostio jaz između općeg potencijala temeljnog modela i zahtjeva praktične primjene, nužna je daljnja faza prilagodbe koja se tipično sastoji od dva komplementarna procesa: **finog podešavanja (fine-tuning)** za specijalizaciju zadataka i **poravnanja (alignment)** za oblikovanje ponašanja.

**Fino podešavanje** predstavlja primjenu **transfer learninga** (prijenosu učenja), moćne paradigme u strojnom učenju gdje se znanje stečeno rješavanjem jednog problema (u ovom slučaju, opće modeliranje jezika tijekom pre-treniranja) koristi kao polazišna točka za rješavanje drugog, srodnog problema (Howard & Ruder, 2018). Umjesto treniranja modela od nule za svaki specifični zadatak, što bi bilo izuzetno neefikasno s obzirom na veličinu LLM-ova, fino podešavanje započinje s parametrima već pre-treniranog modela. Model se zatim dodatno trenira, obično koristeći standardne tehnike nadziranog učenja, na znatno manjem, pažljivo kuriranim skupu podataka koji je specifičan za ciljni zadatak. Ovi podaci sastoje se od primjera ulaza i željenih izlaza (oznaka) za taj zadatak – na primjer, skup parova pitanje-odgovor za razvoj chatbota, zbirka medicinskih tekstova s anotacijama za primjenu u zdravstvu, repozitoriji programskog koda s komentarima za alate za pomoć pri kodiranju, ili parovi izvornih rečenica i njihovih prijevoda za strojno prevodenje. Tijekom ovog procesa, parametri modela (ili barem neki njihov dio) se lagano ažuriraju optimizacijom standardne funkcije gubitka (npr. unakrsna entropija) na specifičnom skupu podataka. Cilj je specijalizirati model za nijanse i zahtjeve ciljnog zadatka ili domene, istovremeno zadržavajući i iskorištavajući bogato opće lingvističko i činjenično znanje koje je model stekao tijekom pre-treniranja. Ovaj pristup značajno smanjuje potrebu za velikim količinama označenih podataka za svaki novi zadatak i ubrzava razvoj specijaliziranih AI aplikacija. Međutim, fino podešavanje također nosi rizik od **katastrofalnog zaboravljanja (catastrophic forgetting)** (McCloskey & Cohen, 1989), gdje model tijekom prilagodbe novom zadatku može izgubiti dio općih sposobnosti naučenih tijekom pre-treniranja, što zahtijeva pažljive strategije treniranja (npr. niske stope učenja, postupno "odmrzavanje" slojeva).

Dok fino podešavanje primarno prilagodava *sposobnosti* modela za specifične zadatke, **poravnanje (alignment)** fokusira se na oblikovanje njegovog *ponašanja*. Čak i model koji je fino podešen za određeni zadatak i tehnički sposoban generirati relevantne odgovore, može i dalje proizvoditi sadržaj koji je nepoželjan – na primjer, može biti nekoristan, davati netočne ili izmišljene informacije (halucinirati), generirati pristran ili toksičan tekst, ili slijediti štetne upute. Poravnanje je stoga ključan i složen proces čiji je cilj osigurati da se ponašanje modela što više podudara s ljudskim namjerama i vrijednostima, te da slijedi određene principi, često sažete kao težnja da model bude **koristan (helpful)**, **iskren (honest)** i **bezopasan (harmless)** (Askell et al., 2021). Ovo nije samo tehnički izazov, već i duboko etičko pitanje koje uključuje definiranje željenih normi ponašanja i razvoj metoda za njihovo usađivanje u model.

Dominantna tehnika koja se danas koristi za postizanje poravnjanja jest **učenje s pojačanjem uz ljudsku povratnu informaciju (Reinforcement Learning from Human Feedback - RLHF)**. Ovaj pristup, koji je stekao popularnost kroz radove poput onih iz OpenAI-ja (Stiennon et al.,

2020; Ouyang et al., 2022) na modelima InstructGPT i ChatGPT, koristi ljudske preferencije kao signal za usmjeravanje učenja modela. Proces RLHF obično uključuje nekoliko koraka:

1. **(Opcionalno, ali često) Nadzirano fino podešavanje (Supervised Fine-Tuning - SFT):** Prije samog RLHF-a, pre-trenirani model se često prvo fino podešava na manjem skupu visokokvalitetnih demonstracija željenog ponašanja (npr. skup promptova i odgovora koje su napisali ljudi ili pažljivo filtrirani odgovori samog modela). Ovo daje modelu početnu orientaciju prema željenom stilu i formatu odgovora.
2. **Prikupljanje podataka o ljudskim preferencijama:** Generira se skup promptova (uputa) na koje početni (obično SFT) model daje više različitih odgovora (npr. dva ili više). Ljudi (ocjenjivači, labeleri) zatim uspoređuju te odgovore i odabiru onaj koji preferiraju prema zadanim kriterijima (korisnost, istinitost, bezopasnost, itd.) ili ih rangiraju od najboljeg do najgoreg. Rezultat je skup podataka koji ne sadrži apsolutne "točne" odgovore, već relativne preferencije između različitih mogućih odgovora.
3. **Treniranje modela nagrađivanja (Reward Model - RM):** Na temelju prikupljenih podataka o ljudskim preferencijama, trenira se zaseban model (obično manji LLM ili modificirana verzija glavnog modela) koji uči predviđati koju bi ocjenu (nagradu) ljudi dali određenom paru (prompt, odgovor). Ovaj model uči mapirati odgovor na skalaru vrijednost koja odražava ljudsku preferenciju. Često se koristi pristup temeljen na modelima poput Bradley-Terry modela za učenje iz usporedbi parova (Bradley & Terry, 1952).
4. **Fino podešavanje LLM-a pomoću učenja s pojačanjem (RL):** Glavni LLM (koji je prošao SFT) se dalje fino podešava koristeći algoritme učenja s pojačanjem. U ovoj fazi, LLM djeluje kao **agent** čija je **politika (policy)** generiranje odgovora na promptove. Za svaki generirani odgovor, **model nagrađivanja (RM)** pruža skalarnu **nagradu (reward)**. Cilj RL algoritma (često se koristi Proximal Policy Optimization - PPO (Schulman et al., 2017) zbog svoje relativne stabilnosti i efikasnosti) jest ažurirati parametre LLM-a (njegovu politiku) tako da maksimizira očekivanu nagradu koju dobiva od RM-a. Kako bi se spriječilo da LLM previše odstupi od jezika naučenog tijekom SFT faze i počne generirati besmislice samo da bi "prevario" model nagrađivanja (reward hacking), obično se dodaje regularizacijski član u funkciju cilja RL algoritma, najčešće kao kazna za KL divergenciju između trenutne politike LLM-a i njegove politike nakon SFT faze.

Iako je RLHF postao industrijski standard, on je složen, računalno intenzivan i ovisi o skupom i potencijalno nekonzistentnom procesu prikupljanja ljudskih preferencija. Stoga se aktivno istražuju i razvijaju alternativne i komplementarne tehnike poravnjanja:

- **Direct Preference Optimization (DPO)** (Rafailov et al., 2023): Ovaj pristup nastoji pojednostaviti RLHF izbjegavajući eksplicitno treniranje zasebnog modela nagrađivanja i korištenje RL algoritama. DPO izvodi analitičku vezu između optimalnog modela nagrađivanja i optimalne politike te formulira funkciju gubitka koja omogućuje direktno fino podešavanje LLM-a na podacima o ljudskim preferencijama koristeći samo standardno nadzirano učenje (binarnu unakrsnu entropiju). Pokazalo se da DPO može postići slične ili čak bolje performanse od RLHF-a uz znatno manju složenost implementacije i treniranja.
- **Reinforcement Learning from AI Feedback (RLAIF)** (Bai et al., 2022): Motivirani izazovima skaliranja prikupljanja ljudskih povratnih informacija, RLAIF predlaže

korištenje drugog AI modela (često snažnog LLM-a) za pružanje preferencijskih oznaka umjesto ljudi. Proces uključuje definiranje skupa principa ili "ustava" (Constitution) prema kojem AI ocjenjivač treba vrednovati odgovore (npr. "Budi koristan", "Nemoj generirati ilegalni sadržaj"). Glavni LLM se zatim poravnava koristeći ove AI-generirane preferencije, potencijalno kroz RL ili DPO. Ovaj pristup, poznat i kao "Constitutional AI", obećava veću skalabilnost i konzistentnost, ali nosi rizik da AI modeli pojačavaju vlastite pristranosti i ograničenja, te postavlja pitanje kako definirati adekvatan "ustav".

- **Robust Alignment Fine-Tuning (RAFT)** (Dong et al., 2023; primjer naziva, specifične tehnike za robusnost se stalno razvijaju): Prepoznajući da poravnanje postignuto na određenom skupu podataka možda neće biti robusno na varijacije u promptovima ili na pokušaje zlonamjernih napada, razvijaju se tehnike koje eksplicitno ciljaju na poboljšanje robusnosti poravnjanja. To može uključivati treniranje na raznolikijim ili čak adversarijski generiranim promptovima, korištenje tehnika regularizacije koje potiču generalizaciju, ili specifične metode finog podešavanja dizajnirane da održe poravnanje čak i pod distribucijskim pomacima.

Poravnanje je, bez sumnje, kritično za izgradnju povjerenja i sigurno uvođenje LLM-ova u široku upotrebu. Ono omogućuje smanjenje generiranja štetnog, pristranog ili netočnog sadržaja i povećava korisnost modela u interakciji s ljudima. Međutim, proces poravnjanja nije bez izazova. Definiranje univerzalno prihvatljivih ljudskih vrijednosti i preferencija je inherentno težak, ako ne i nemoguć zadatak, podložan kulturnim i individualnim razlikama. Vrijednosti i preferencije ljudskih ocjenjivača koji pružaju povratne informacije, ili principi ugrađeni u ugradbene pristupe (Bai et al., 2022), mogu nenamjerno uvesti nove slojeve pristranosti ili odražavati specifične kulturne norme. Osiguravanje pravednog i reprezentativnog poravnjanja za raznolike globalne korisnike s kojima će agenci komunicirati ostaje ključno otvoreno pitanje (Casper et al., 2023). Osim toga, postoji i fenomen poznat kao "**alignment tax**" (porez na poravnanje), gdje proces poravnjanja, iako poboljšava sigurnost i korisnost u razgovornom smislu, može istovremeno blago smanjiti performanse modela na nekim standardnim akademskim benchmarkovima ili ograničiti njegovu kreativnost (Askell et al., 2021). A i osiguravanje robusnosti poravnjanja protiv namjernih pokušaja zaobilaženja (jailbreaking) izgleda poput neprestane "igre mačke i miša".

### 3.3.3 Optimizacija za Stvarni Svet: Efikasnost (Efficiency)

Sama veličina koja karakterizira velike jezične modele, premda ključna za njihove napredne sposobnosti, istovremeno predstavlja i njihovu Ahilovu petu u kontekstu praktične primjene. Kako modeli rastu do stotina milijardi ili čak trilijuna parametara, njihovo treniranje i, što je još važnije za krajnje korisnike, njihovo pokretanje u fazi **inferencije** (generiranja odgovora u stvarnom vremenu) postaju izuzetno zahtjevni. Ovi procesi gutaju ogromne količine specijaliziranih računalnih resursa, poput grafičkih procesorskih jedinica (GPU) ili tenzorskih procesorskih jedinica (TPU), zahtijevaju značajne količine memorije (kako za pohranu samih parametara modela, tako i za privremene aktivacije tijekom izračuna) i troše velike količine električne energije, što povlači za sobom ne samo visoke operativne troškove već i značajan okolišni otisak (Strubell et al., 2019; Patterson et al., 2021; Luccioni et al., 2022). Ova resursna intenzivnost predstavlja značajnu prepreku širokoj dostupnosti i implementaciji LLM-ova, ograničavajući njihovu upotrebu na organizacije s dubokim džepovima i snažnom infrastrukturom, te otežavajući njihovo pokretanje na uređajima s ograničenim resursima, poput pametnih telefona ili rubnih (edge) uređaja. Stoga je **optimizacija efikasnosti** postala kritično područje istraživanja i razvoja, usmjereno na smanjenje računalnih, memorijskih i energetskih zahtjeva LLM-ova bez značajnog žrtvovanja njihovih performansi, čime se otvara

put njihovoj demokratizaciji i održivijoj primjeni. Nekoliko ključnih strategija dominira ovim područjem: parametarski efikasno fino podešavanje, kvantizacija i pruning.

Jedan od prvih izazova efikasnosti javlja se tijekom prilagodbe pre-treniranih modela specifičnim zadacima. Tradicionalno **puno fino podešavanje (full fine-tuning)** uključuje ažuriranje *svih* parametara pre-treniranog modela na novom, specifičnom skupu podataka. Iako efikasno u smislu postizanja dobrih performansi, ovo stvara ozbiljan problem skalabilnosti: za svaki novi zadatak potrebno je pohraniti potpunu, ogromnu kopiju fino podešenog modela, što postaje izuzetno skupo u smislu pohrane i upravljanja ako se model treba prilagoditi za desetke ili stotine različitih zadataka. Kako bi se riješio ovaj problem, razvijene su tehnike **parametarski efikasnog finog podešavanja (Parameter-Efficient Fine-Tuning - PEFT)**.

Osnovna ideja PEFT metoda jest zamrznuti veliku većinu (ili sve) parametre originalnog pre-treniranog modela i ažurirati samo mali broj dodatnih ili postojećih parametara tijekom finog podešavanja (Lialin et al., 2023; He et al., 2021). Broj parametara koji se treniraju često je manji od 1% ukupnog broja parametara modela, što drastično smanjuje memorijske zahtjeve (jer treba pohraniti samo male skupove dodatnih parametara za svaki zadatak, a ne cijeli model) i računalne troškove finog podešavanja. Nekoliko popularnih PEFT pristupa uključuje:

- **Adapteri (Adapters):** Ova tehnika umeće male, dodatne neuronske module, nazvane adapteri, unutar svakog sloja (ili nekih slojeva) originalne Transformer arhitekture (Houlsby et al., 2019; Pfeiffer et al., 2020). Adapteri obično imaju arhitekturu uskog grla (bottleneck) – prvo projiciraju ulaznu aktivaciju u prostor niže dimenzije, primjenjuju nelinearnu funkciju, a zatim je projiciraju natrag u originalnu dimenziju. Tijekom finog podešavanja, samo se parametri ovih malih adaptera treniraju, dok su svi parametri originalnog LLM-a zamrznuti. Ovo omogućuje veliku modularnost – za svaki novi zadatak trenira se samo novi set adaptera.
- **LoRA (Low-Rank Adaptation):** Ova iznimno popularna tehnika temelji se na hipotezi da promjene težina tijekom adaptacije modela na specifični zadatak imaju nisku "intrinzičnu rangiranost" (Hu et al., 2021). Umjesto direktnog ažuriranja originalne matrice težina  $W$  (koja može biti vrlo velika), LoRA uvodи dvije manje matrice,  $A$  i  $B$ , tako da je promjena težine aproksimirana njihovim produktom niske rangiranosti:  $\Delta W = BA$ . Tijekom finog podešavanja treniraju se samo matrice  $A$  i  $B$ , čiji je ukupan broj parametara znatno manji od broja parametara u  $W$ . Prilikom inferencije, promjena  $BA$  se jednostavno doda originalnoj težini  $W$ . LoRA se pokazala vrlo efikasnom i postiže performanse usporedive s punim finim podešavanjem uz djelić treniranih parametara.
- **Podešavanje prompta (Prompt Tuning) i Prefiksa (Prefix Tuning):** Ove metode zauzimaju drugačiji pristup. Umjesto modificiranja internih težina modela, one dodaju male skupove kontinuiranih vektora (koji se uče) na ulaz modela (Lester et al., 2021) ili u skrivene slojeve (Li & Liang, 2021). Ovi naučeni vektori, zvani "soft prompts" ili "prefixes", djeluju kao specifične upute koje usmjeravaju ponašanje zamrznutog pre-treniranog modela prema željenom zadatku, bez potrebe za mijenjanjem jednog od njegovih originalnih parametara. Ovo je konceptualno slično diskretnom prompt inženjeringu, ali s prednošću što se optimalni "prompt" uči automatski kroz gradijentni spust.

PEFT tehnike revolucionirale su način na koji se LLM-ovi prilagođavaju, čineći proces znatno bržim, memorijski efikasnijim i omogućujući lakše upravljanje i posluživanje više različitih zadataka koristeći isti temeljni model.

Druga ključna strategija za optimizaciju efikasnosti, primjenjiva i tijekom treniranja i, još važnije, tijekom inferencije, jest **kvantizacija (Quantization)**. Ona se odnosi na proces smanjenja numeričke preciznosti korištene za predstavljanje parametara (težina) modela i/ili njegovih internih aktivacija tijekom izračuna. Standardni modeli obično koriste 32-bitne (FP32) ili 16-bitne (FP16 ili BF16) brojeve s pomičnim zarezom. Kvantizacija ima za cilj pretvoriti ove vrijednosti u formate niže preciznosti, najčešće 8-bitne cijele brojeve (INT8), a u novije vrijeme čak i 4-bitne (INT4) ili još niže formate (Dettmers et al., 2022; Frantar et al., 2022; Yao et al., 2022). Glavne prednosti kvantizacije su višestruke: prvo, drastično smanjuje **memorijski otisak** modela (npr. prelazak s FP16 na INT8 prepovoljuje veličinu modela, a na INT4 je smanjuje za četiri puta), što omogućuje pokretanje većih modela na hardveru s ograničenom memorijom; drugo, operacije s nižom preciznošću (posebice cjelobrojne) mogu biti značajno **brže** na hardveru koji ih podržava (mnogi moderni procesori i akceleratori imaju specijalizirane jedinice za ubrzanje INT8 operacija); treće, smanjeni prijenos podataka i jednostavnije operacije mogu dovesti do **manje potrošnje energije**. Glavni izazov kvantizacije jest očuvanje točnosti modela, jer smanjenje preciznosti neizbjegno uvodi određenu pogrešku. Postoje dvije glavne paradigme: **Post-Training Quantization (PTQ)**, gdje se već pre-trenirani model kvantizira bez dodatnog treniranja (često koristeći mali kalibracijski skup podataka za određivanje optimalnih parametara kvantizacije), što je brže ali može dovesti do većeg pada točnosti; i **Quantization-Aware Training (QAT)**, gdje se efekt kvantizacije simulira tijekom procesa finog podešavanja ili čak pre-treniranja, omogućujući modelu da se prilagodi nižoj preciznosti i tako zadrži višu razinu točnosti (Jacob et al., 2018). Zahvaljujući napretku u kvantizacijskim tehnikama, danas je često moguće postići značajnu kompresiju (npr. na 4 bita) uz vrlo mali ili gotovo nikakav gubitak performansi na nizvodnim zadacima, čineći kvantizaciju izuzetno popularnom metodom za efikasnu implementaciju LLM-ova.

Treća važna tehnika optimizacije je **pruning (obrezivanje)**, koja se temelji na ideji da su mnogi veliki neuronski modeli **preparametrizirani**, odnosno sadrže značajan broj parametara (težina) ili čak cijelih strukturalnih komponenti koje su redundantne ili ne doprinose značajno končnim performansama modela. Pruning ima za cilj identificirati i trajno ukloniti te manje važne elemente, čime se smanjuje veličina modela i potencijalno ubrzava inferencija (LeCun et al., 1990; Han et al., 2015). Razlikujemo dva glavna tipa pruninga:

- **Nestrukturirani (Unstructured) Pruning:** Ova metoda uklanja pojedinačne težine iz matrica težina modela, obično na temelju nekog kriterija važnosti (npr. uklanjanje težina s najmanjom apsolutnom vrijednošću - magnitude pruning). Iako može postići vrlo visoke stope prorjeđivanja (npr. ukloniti 90% ili više težina) uz zadržavanje točnosti, rezultirajuće matrice težina postaju rijetke (sparse) s nepravilnim uzorcima nula. Standardni hardver (GPU, CPU) često nije optimiziran za efikasno iskorištavanje ovakve fine granularne rijetkosti, pa postignuto smanjenje broja parametara ne mora nužno dovesti do značajnog ubrzanja inferencije u praksi, osim ako se ne koristi specijalizirani hardver ili softverske biblioteke.
- **Strukturirani (Structured) Pruning:** Ova metoda uklanja cijele, pravilne strukturalne jedinice modela, kao što su pojedinačni neuroni, kanali u konvolucijskim mrežama, ili, u kontekstu Transformer-a, cijele glave pažnje (attention heads) ili čak cijeli slojevi (Li et al., 2016; Voita et al., 2019; Michel et al., 2019). Budući da se uklanjuju cijeli blokovi, rezultirajući model je manji, ali i dalje ima gustu strukturu koja se može efikasno izvršavati na standardnom hardveru, često dovodeći do stvarnih ubrzanja inferencije uz

smanjenje broja parametara. Određivanje koje strukture ukloniti obično se temelji na procjeni njihove važnosti za performanse modela.

Pruning se često izvodi iterativno: model se trenira, dio parametara se obreže, a zatim se model dodatno fino podešava kako bi se oporavio od potencijalnog pada točnosti uzrokovanih obrezivanjem. Zanimljiva poveznica je **Hipoteza lutrijskog listića (Lottery Ticket Hypothesis)** (Frankle & Carbin, 2019), koja postulira da unutar velikih, nasumično inicijaliziranih neuronskih mreža postoje male pod-mreže ("dubitni listići") koje, kada se treniraju u izolaciji od samog početka s istom inicijalizacijom, mogu postići performanse usporedive s originalnom velikom mrežom. Iako pronađenje ovih "listića" nije trivijalno, ova hipoteza dodatno podupire ideju da su veliki modeli inherentno redundantni i da postoje znatno manje, efikasnije arhitekture koje čekaju da budu otkrivene ili ekstrahirane. Tehnike poput **destilacije znanja (knowledge distillation)** (Hinton et al., 2015), gdje se znanje velikog "učiteljskog" modela prenosi na manji "učenički" model, također se često koriste, ponekad u kombinaciji s pruningom, za stvaranje manjih i bržih modela (npr. DistilBERT (Sanh et al., 2020)).

Tehnike poput parametarski efikasnog finog podešavanja, kvantizacije i pruninga, često korištene i u kombinaciji, predstavljaju ključne alete za premoćivanje jaza između teorijskih sposobnosti velikih jezičnih modela i zahtjeva praktične primjene u stvarnom svijetu. One čine LLM-ove ne samo izvedivijima za implementaciju u širem rasponu scenarija, uključujući one s ograničenim resursima, već i smanjuju njihove ekonomski i ekološke troškove, doprinoseći njihovoj demokratizaciji i otvarajući put prema održivoj budućnosti umjetne inteligencije.

### 3.3.4 Mjerenje Uspjeha: Evaluacija (Evaluation)

Nakon što je veliki jezični model prošao kroz faze pre-treniranja, finog podešavanja i potencijalno poravnjana, postavlja se ključno pitanje: kako objektivno procijeniti njegovu kvalitetu i sposobnosti? **Evaluacija** velikih jezičnih modela predstavlja fundamentalan, ali istovremeno i izuzetno izazovan korak u njihovom životnom ciklusu. Ona nije samo akademska vježba za usporedbu različitih modela ili tehnika, već je presudna za razumijevanje stvarnih mogućnosti i ograničenja ovih sustava, za usmjeravanje daljnog razvoja, te za donošenje informiranih odluka o njihovoj primjeni u stvarnom svijetu. Izazov proizlazi iz same prirode jezika – njegove inherentne višežnačnosti, kontekstualne ovisnosti i suptilnosti – te iz činjenice da "dobar" jezični model može značiti različite stvari ovisno o zadatku i krajnjem korisniku. Ne postoji jedna jedinstvena, savršena metrika ili benchmark koji bi mogao sveobuhvatno ocijeniti sve aspekte performansi LLM-a. Stoga se evaluacija obično provodi kroz niz različitih zadataka i metrika, često podijeljenih u dvije široke kategorije: procjenu sposobnosti **razumijevanja prirodnog jezika (Natural Language Understanding - NLU)** i procjenu sposobnosti **generiranja prirodnog jezika (Natural Language Generation - NLG)**, uz sve veći naglasak na holističke pristupe koji pokušavaju obuhvatiti i druge važne dimenzije poput robustnosti, pravednosti i efikasnosti.

Evaluacija **razumijevanja prirodnog jezika (NLU)** usmjerena je na procjenu sposobnosti modela da interpretira, analizira i izvuče značenje iz danog tekstualnog ulaza. Ovo se tipično postiže testiranjem modela na standardiziranim skupovima podataka (benchmarkovima) koji pokrivaju različite temeljne NLU zadatke. Svaki zadatak dizajniran je da ispita specifičan aspekt jezičnog razumijevanja, a performanse se mjere kvantitativnim metrikama uspoređivanjem izlaza modela s unaprijed definiranim "točnim" odgovorima (ground truth). Neki od ključnih NLU zadataka i primjeri benchmarka uključuju:

- **Klasifikacija teksta:** Ovo je jedan od najosnovnijih NLU zadataka, gdje je cilj dodijeliti unaprijed definiranu kategoriju ili oznaku cijelom tekstu. Primjeri uključuju **analizu sentimenta**, gdje se tekst klasificira kao pozitivan, negativan ili neutralan (npr. koristeći skupove podataka poput IMDb recenzija filmova (Maas et al., 2011) ili SST-2 (Socher et al., 2013)), ili **klasifikaciju tema**, gdje se novinski članci svrstavaju u kategorije poput sporta, politike ili tehnologije (npr. AG News dataset). Uobičajene metrike za evaluaciju su **točnost (Accuracy)**, koja mjeri postotak ispravno klasificiranih primjera, i **F1-skor**, koji predstavlja harmonijsku sredinu preciznosti (precision) i odziva (recall) te je posebno koristan kod neuravnoteženih skupova podataka gdje jedna kategorija dominira.
- **Prepoznavanje imenovanih entiteta (Named Entity Recognition - NER):** Cilj ovog zadatka je identificirati i klasificirati spomene imenovanih entiteta unutar teksta u predefinirane kategorije kao što su osobe, organizacije, lokacije, datumi, itd. Na primjer, u rečenici "Apple je objavio nove rezultate iz Cupertino", NER sustav bi trebao identificirati "Apple" kao organizaciju i "Cupertino" kao lokaciju. Klasični benchmark za ovaj zadatak je CoNLL-2003 (Tjong Kim Sang & De Meulder, 2003). Budući da se radi o zadatku označavanja sekvenci (svaki token dobiva oznaku), performanse se najčešće mjeru **F1-skorom** izračunatim na razini entiteta.
- **Odgovaranje na pitanja (Question Answering - QA):** Ovaj zadatak procjenjuje sposobnost modela da odgovori na pitanja postavljena u prirodnom jeziku. Postoji više varijanti: **ekstraktivno QA**, gdje model mora pronaći odgovor kao kontinuirani segment (span) teksta unutar zadanoj kontekstualne odlomke (npr. SQuAD - Stanford Question Answering Dataset (Rajpurkar et al., 2016, 2018))), **apstraktivno QA**, gdje model mora generirati odgovor svojim riječima, potencijalno sintetizirajući informacije iz više dijelova konteksta ili iz svog internog znanja, i **otvoreno-domensko QA**, gdje model mora odgovoriti na pitanja o općem znanju bez eksplicitno zadanoj konteksta (npr. Natural Questions (Kwiatkowski et al., 2019), TriviaQA (Joshi et al., 2017)). Za ekstraktivno QA, uobičajene metrike su **Exact Match (EM)**, postotak odgovora koji se točno poklapaju s točnim odgovorom, i **F1-skor**, koji mjeri prosječno preklapanje tokena između predviđenog i točnog odgovora. Evaluacija apstraktivnog i otvoreno-domenskog QA znatno je izazovnija i često se oslanja na metrike slične onima za NLG (poput BLEU ili ROUGE) ili ljudsku procjenu.
- **Zaključivanje u prirodnom jeziku (Natural Language Inference - NLI):** Poznat i kao prepoznavanje tekstualnog slijeda (Recognizing Textual Entailment - RTE), ovaj zadatak procjenjuje sposobnost modela da odredi logički odnos između dva teksta, obično nazvanih premisa i hipoteza. Odnos može biti **slijed (entailment)** (hipoteza logički slijedi iz premise), **kontradikcija (contradiction)** (hipoteza proturječi premisi), ili **neutralan (neutral)** (nema jasnog logičkog odnosa). Primjeri benchmarka uključuju SNLI (Stanford Natural Language Inference) (Bowman et al., 2015) i MNLI (Multi-Genre NLI) (Williams et al., 2018). Standardna metrika je **točnost** klasifikacije odnosa.

Kako bi se omogućila standardizirana i sveobuhvatnija evaluacija NLU sposobnosti, razvijeni su **agregativni benchmarkovi** koji kombiniraju više različitih zadataka i skupova podataka. Najpoznatiji primjeri su **GLUE (General Language Understanding Evaluation)** (Wang et al., 2018) i njegov zahtjevniji nasljednik **SuperGLUE** (Wang et al., 2019). Ovi benchmarkovi pružili su važan okvir za usporedbu napretka različitih modela. Međutim, kako su LLM-ovi postajali sve moćniji, performanse na ovim benchmarkovima brzo su dosegle ili čak premašile procijenjene

ljudske performanse, što je dovelo do zabrinutosti oko njihove zasićenosti i stvarne sposobnosti razlikovanja modela (tzv. "benchmark saturation"). Stoga su razvijeni novi, izazovniji benchmarkovi poput **MMLU (Massive Multitask Language Understanding)** (Hendrycks et al., 2020), koji procjenjuje znanje i sposobnost rezoniranja modela na širokom spektru od 57 akademskih i stručnih područja (od matematike i fizike do prava i povijesti), težeći mjerenu općenitije inteligencije. Unatoč napretku, ostaje izazov osigurati da performanse na ovim statičkim benchmarkovima doista odražavaju sposobnost modela u dinamičnim i nepredvidljivim uvjetima stvarnog svijeta.

Evaluacija **generiranja prirodnog jezika (NLG)**, s druge strane, predstavlja znatno veći izazov zbog inherentne subjektivnosti onoga što čini "dobar" generirani tekst. Za razliku od NLU zadataka gdje često postoji jedan točan odgovor, kod NLG zadataka poput sažimanja, prevodenja ili kreativnog pisanja, može postojati mnogo jednakih valjanih ili dobrih izlaza. Stoga se evaluacija NLG-a oslanja na različite pristupe, svaki sa svojim prednostima i nedostacima:

- **Metrike temeljene na preklapanju (Overlap-based Metrics):** Ove metrike pokušavaju kvantificirati kvalitetu generiranog teksta mjeranjem njegove sličnosti s jednim ili više referentnih tekstova (koje su napisali ljudi). Najpoznatije metrike u ovoj kategoriji su **BLEU (Bilingual Evaluation Understudy)** (Papineni et al., 2002), primarno korištena u strojnem prevodenju, koja mjeri preciznost preklapanja n-grama (sekvenci od n riječi) između generiranog prijevoda i skupa referentnih prijevoda, uz dodatak kazne za prekratke prijevode (brevity penalty); i **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)** (Lin, 2004), često korištena za evaluaciju sažimanja, koja mjeri odziv (recall) preklapanja n-grama (ROUGE-N) ili najduže zajedničke podsekvence (ROUGE-L) između generiranog sažetka i referentnih sažetaka. Prednost ovih metrika je što su brze, jeftine za izračun i pružaju standardiziranu mjeru. Međutim, njihov glavni nedostatak je često **slaba korelacija s ljudskim prosudbama** o kvaliteti, posebno u pogledu aspekata kao što su fluentnost, koherentnost, adekvatnost značenja i činjenična točnost. One su vrlo osjetljive na specifičan izbor riječi i mogu nepravedno kazniti generirani tekst koji je semantički ispravan i fluentan, ali koristi drugačije riječi ili strukturu rečenice od referentnih tekstova (npr. dobra parafraza može dobiti nizak BLEU/ROUGE skor).
- **Metrike temeljene na embeddingsima (Embedding-based Metrics):** Kako bi se prevladala ograničenja površinskog preklapanja n-grama, razvijene su metrike koje koriste vektorske reprezentacije (embeddings) riječi ili rečenica za mjerjenje semantičke sličnosti između generiranog i referentnog teksta. Primjer je **BERTScore** (Zhang et al., 2019), koji koristi kontekstualizirane embeddinge iz BERT modela za izračun sličnosti između svakog tokena u generiranom tekstu i svakog tokena u referentnom tekstu, a zatim agregira te sličnosti kako bi dobio konačnu ocjenu. Druge slične metrike uključuju MoverScore (Zhao et al., 2019). Ove metrike generalno pokazuju bolju korelaciju s ljudskim prosudbama o semantičkoj sličnosti nego metrike temeljene na preklapanju. Ipak, i one se oslanjaju na postojanje referentnih tekstova i možda neće adekvatno uhvatiti druge aspekte kvalitete poput činjenične točnosti ili koherentnosti na razini diskursa.
- **Metrike temeljene na modelima (Model-based Metrics):** S porastom moći samih LLM-ova, pojavio se novi trend korištenja jednog (ili više) LLM-a kao automatskog evaluatorsa za procjenu kvalitete teksta generiranog od strane drugog LLM-a. Ovi pristupi, ponekad nazivani "AI kritičari", mogu biti dizajnirani da procjenjuju specifične dimenzije

kvalitete, poput koherentnosti, relevantnosti, fluentnosti, činjenične točnosti ili čak stila, često bez potrebe za referentnim tekstovima (npr. G-Eval (Liu et al., 2023b), GPTScore (Fu et al., 2023)). Potencijal ovih metoda je velik, jer bi mogle pružiti nijansiranju i semantički bogatiju evaluaciju. Međutim, one također nose rizike: evaluacija može biti pod utjecajem pristranosti samog evaluator modela, postoji opasnost od "samopogačavanja" gdje modeli favoriziraju izlaze slične vlastitim, a trošak korištenja velikih modela za evaluaciju može biti značajan. Pouzdanost i kalibracija ovih model-baziranih metrika aktivno su područje istraživanja.

- **Ljudska evaluacija:** Unatoč razvoju automatskih metrika, **ljudska prosudba** se i dalje smatra "zlatnim standardom" za evaluaciju NLG-a, kao i za mnoge aspekte NLU-a i općenito ponašanja LLM-a. Ljudi mogu procijeniti suptilne nijanske kvalitete koje automatske metrike često propuštaju. Evaluacija se obično provodi tako da ljudi (anotatori, ocjenjivači) ocjenjuju generirani tekst prema nizu unaprijed definiranih kriterija, kao što su **fluentnost** (gramatička ispravnost i prirodnost jezika), **koherentnost** (logička povezanost i smislenost teksta), **relevantnost** (koliko dobro odgovor odgovara na prompt ili zadatak), **činjenična točnost** (usklađenost s poznatim činjenicama), **korisnost, bezopasnost**, itd. Ocjene se mogu davati na Likertovim skalamama (npr. od 1 do 5), kroz usporedbu parova (koji je od dva odgovora bolji?) ili rangiranjem više odgovora. Glavni nedostaci ljudske evaluacije su što je izuzetno **skupa, vremenski zahtjevna, teško skalabilna** i podložna **subjektivnosti i nekonzistentnosti** (različiti ocjenjivači mogu imati različite interpretacije kriterija ili različite preferencije). Osiguravanje kvalitete i pouzdanosti ljudske evaluacije zahtijeva pažljiv dizajn zadatka, jasne smjernice i obuku ocjenjivača.

Prepoznujući ograničenja evaluacije temeljene na pojedinačnim zadacima ili metrikama, sve je veći naglasak na **holističkoj evaluaciji**. Cilj je procijeniti LLM-ove na širem spektru sposobnosti i karakteristika, često kroz kombinaciju različitih zadataka, scenarija i metrika. Jedan od najambicioznijih napora u ovom smjeru je **HELM (Holistic Evaluation of Language Models)** (Liang et al., 2022) sa Sveučilišta Stanford. HELM nastoji pružiti višedimenzionalnu sliku performansi modela evaluirajući ih na desetinama scenarija (koji pokrivaju različite zadatke i domene) i mjereći ne samo standardnu **točnost**, već i druge ključne atribute kao što su **kalibracija** (koliko su pouzdane procjene vjerojatnosti modela), **robustnost** (performanse pod perturbacijama ulaza ili distribucijskim pomacima), **pravednost** (odsutnost neželjenih pristranosti prema različitim demografskim skupinama), **bias** (npr. društveni stereotipi), **toksičnost** (sklonost generiranju štetnog sadržaja) i **efikasnost** (računalni resursi potrebni za inferenciju). Drugi važni alati i okviri za širu evaluaciju uključuju EleutherAI-jev LM Evaluation Harness. Nadalje, raste interes za evaluaciju specifičnih, naprednijih sposobnosti poput **rezoniranja** (npr. na matematičkim problemima kao u GSM8K (Cobbe et al., 2021) ili logičkim zagonetkama), **činjenične točnosti i istinitosti** (npr. TruthfulQA (Lin et al., 2021), koji mjeri sklonost modela generiraju uobičajenih zabluda), te **sigurnosti i otpornosti** na **zlonamjerne napade** (adversarial robustness).

Zaključno, evaluacija velikih jezičnih modela je složen, višeslojan i dinamičan proces koji je daleko od riješenog problema. Dok standardizirani benchmarkovi i automatske metrike pružaju korisne, kvantitativne pokazatelje napretka, oni često ne uspijevaju uhvatiti punu sliku performansi modela u stvarnom svijetu i mogu biti podložni "iskorištanju" (gaming). Ljudska evaluacija ostaje nezamjenjiva za procjenu nijansiranih aspekata kvalitete, ali je skupa i teško skalabilna. Holistički pristupi koji kombiniraju različite zadatke, metrike i aspekte (uključujući

etičke) obećavaju pružanje sveobuhvatnije slike, ali zahtijevaju značajne resurse i kontinuirano ažuriranje kako bi pratili brzi napredak samih modela. Kako LLM-ovi postaju sve sposobniji i integriraniji u naše živote, razvoj rigoroznih, pouzdanih i smislenih metoda evaluacije postaje ne samo znanstveni izazov, već i društvena nužnost za osiguravanje njihovog odgovornog razvoja i primjene.

### 3.3.5 Model u Akciji: Inferencija (Inference)

Faza **inferencije** predstavlja kulminaciju cijelokupnog životnog ciklusa velikog jezičnog modela; to je trenutak kada se sva uložena energija u prikupljanje podataka, pre-treniranje, fino podešavanje, poravnanje i optimizaciju pretvara u oplaplju funkcionalnost i vrijednost za krajnjeg korisnika. Inferencija se odnosi na proces korištenja već istreniranog i prilagođenog modela za obavljanje specifičnih zadataka u stvarnom vremenu – bilo da se radi o generiranju odgovora na korisnički upit, prevođenju teksta, pisanju sažetaka, kreiranju programskog koda ili bilo kojoj drugoj primjeni za koju je model dizajniran. Za razliku od faza treniranja i finog podešavanja koje se mogu odvijati offline i trajati danima ili tjednima, inferencija se tipično događa na zahtjev, često u interaktivnim scenarijima gdje su korisnička očekivanja usmjerena na brže i kvalitetne rezultate. Stoga su dva ključna, često suprotstavljena cilja koja dominiraju dizajnom i optimizacijom inferencijskih sustava: **brzina**, odnosno **niska latencija** (vrijeme potrebno da model generira odgovor), i **kvaliteta izlaza** (relevantnost, točnost, koherencija, fluentnost i sigurnost generiranog teksta). Inferencija je, u suštini, "trenutak istine" gdje se apstraktne sposobnosti modela manifestiraju u konkretnoj interakciji.

Temeljni način interakcije s velikim jezičnim modelom tijekom inferencije odvija se putem **prompta** (upute). Prompt je ulazni tekst koji korisnik (ili drugi sustav) pruža modelu kako bi specificirao zadatak ili pokrenuo generiranje željenog izlaza. Moć i fleksibilnost modernih LLM-ova, posebice onih treniranih na ogromnim i raznolikim skupovima podataka, leži u njihovoj sposobnosti da razumiju i odgovore na upute izražene u prirodnom jeziku. Međutim, pokazalo se da način na koji je prompt formuliran može imati presudan utjecaj na kvalitetu, točnost i relevantnost odgovora koji model generira. Ovo je dovelo do pojave nove discipline, ili barem vještine, poznate kao "**prompt inženjering**" (**prompt engineering**) – umijeće pažljivog dizajniranja i formuliranja efikasnih promptova kako bi se iz modela izvuklo željeno ponašanje i postigle optimalne performanse za određeni zadatak (Liu et al., 2023a; White et al., 2023). Prompt inženjering može uključivati iterativno eksperimentiranje s različitim formulacijama, dodavanje konteksta, specificiranje formata izlaza ili korištenje naprednijih tehnika promptiranja. Nekoliko osnovnih strategija promptiranja su ključne:

- **Zero-shot Prompting:** Ovo je najjednostavniji oblik, gdje se modelu daje samo opis zadataka ili pitanje, bez ikakvih primjera kako bi trebao izgledati željeni izlaz. Na primjer, prompt "Prevedi sljedeću rečenicu s engleskog na francuski: 'Hello, world!'" bio bi zero-shot prompt. Sposobnost LLM-ova da uspješno izvršavaju zadatke u zero-shot postavkama, oslanjajući se isključivo na znanje stečeno tijekom pre-treniranja, jedna je od njihovih najimpresivnijih karakteristika i pokazatelj njihove generalizacijske moći. Međutim, performanse na složenijim ili manje uobičajenim zadacima mogu biti ograničene u ovoj postavci.
- **Few-shot Prompting:** Ova tehnika, popularizirana s GPT-3 modelom (Brown et al., 2020), uključuje pružanje modela ne samo opisa zadataka, već i nekoliko (obično 1 do ~32) primjera parova ulaz-izlaz koji ilustriraju kako zadatak treba riješiti. Ovi primjeri se daju direktno unutar prompta, kao dio konteksta. Na primjer, za zadatak analize

sentimenta, few-shot prompt bi mogao izgledati ovako: "Recenzija: 'Ovaj film je bio užasan.' Sentiment: Negativan\nRecenzija: 'Apsolutno remek-djelo!' Sentiment: Pozitivan\nRecenzija: 'Bilo je ok, ništa posebno.' Sentiment: Neutralan\nRecenzija: 'Gluma je bila fantastična!' Sentiment:". Model zatim koristi ove primjere kao vodič (tzv. **učenje u kontekstu - in-context learning**) kako bi bolje razumio format i prirodu zadatka te generirao odgovor za posljednji, nepotpuni primjer. Few-shot prompting često značajno poboljšava performanse u usporedbi sa zero-shot pristupom, posebice za zadatke koji zahtijevaju specifičan format izlaza ili nijansirano razumijevanje. Važno je napomenuti da tijekom few-shot promptinga ne dolazi do ažuriranja parametara modela; model uči isključivo iz konteksta pruženog u promptu.

- **Chain-of-Thought (CoT) Prompting:** Za zadatke koji zahtijevaju složenje rezoniranje, poput rješavanja matematičkih problema, logičkih zagonetki ili odgovaranja na pitanja koja zahtijevaju višestruke korake zaključivanja, standardni zero-shot ili few-shot promptovi često dovode do netočnih odgovora. Tehnika **Chain-of-Thought (CoT) promptinga** (Wei et al., 2022a) pokazala se izuzetno efikasnom u poboljšanju sposobnosti rezoniranja LLM-ova. Ideja je potaknuti model da eksplicitno generira međukorake ili lanac razmišljanja koji vode do konačnog odgovora, umjesto da samo izbaci konačni rezultat. U few-shot CoT postavkama, primjeri u promptu uključuju ne samo ulaz i konačni izlaz, već i detaljan opis koraka rezoniranja. Na primjer, za matematički problem, primjer bi pokazao kako se problem rastavlja na manje dijelove i rješava korak po korak. Fascinantno je da se slična poboljšanja mogu postići i u zero-shot postavkama jednostavnim dodavanjem fraze poput "Razmislimo korak po korak." (Let's think step by step.) na kraj originalnog prompta (Kojima et al., 2022). Iako mehanizmi iza uspjeha CoT-a nisu potpuno razjašnjeni, pretpostavlja se da eksplicitno generiranje međukoraka alocira više računalnih resursa na problem, omogućuje modelu da prati složenje logičke sljedovite i smanjuje vjerojatnost pogrešaka u rezoniranju.

Nakon što model primi prompt i obradi ga, slijedi faza **generiranja teksta (text generation)** ili **dovršavanja teksta (text completion)**. Budući da su većina modernih generativnih LLM-ova (poput GPT serije) autoregresivni, oni generiraju izlazni tekst sekvencialno, token po token. U svakom koraku generiranja, model uzima u obzir prompt i sve prethodno generirane tokene te izračunava distribuciju vjerojatnosti nad cijelim svojim rječnikom (vocabulary) za sljedeći token. Zadatak **strategije dekodiranja (decoding strategy)** jest odabratiti koji će token iz te distribucije biti sljedeći u generiranoj sekvenci. Izbor strategije dekodiranja ima ključan utjecaj na kvalitetu, koherenciju, raznolikost i determinizam generiranog teksta:

- **Pohlepno pretraživanje (Greedy Search):** Ovo je najjednostavnija strategija. U svakom koraku, ona jednostavno odabire token s najvišom vjerojatnošću prema distribuciji koju daje model. Prednost pohlepnog pretraživanja je njegova brzina i determinizam (za isti ulaz uvijek daje isti izlaz). Međutim, ono često dovodi do suboptimalnih rezultata na razini cijele sekvenca, jer lokalno optimalan izbor u jednom koraku ne garantira globalno optimalnu sekvensu. Pohlepno dekodiranje također ima tendenciju generiranja teksta koji je dosadan, repetitivan ili zapinje u petljama.
- **Pretraživanje snopom (Beam Search):** Kako bi se ublažili problemi pohlepnog pretraživanja, koristi se pretraživanje snopom. Umjesto da prati samo jednu, najvjerojatniju sekvensu, beam search u svakom koraku održava  $k$  (gdje je  $k$  veličina

snopa, "beam width") najvjerojatnijih djelomičnih sekvenci (hipoteza) generiranih do tog trenutka. U sljedećem koraku, za svaku od  $k$  hipoteza generiraju se mogući nastavci za sve tokene u rječniku, te se ponovno odabire  $k$  ukupno najvjerojatnijih sekvenci (obično na temelju sume logaritamskih vjerojatnosti). Ovaj proces se nastavlja dok se ne dosegnu kriteriji za zaustavljanje (npr. generiranje posebnog tokena kraja sekvence ili dosezanje maksimalne duljine). Beam search istražuje veći dio prostora mogućih sekvenci nego pohlepno pretraživanje i često generira tekst više ukupne vjerojatnosti i kvalitete. Međutim, i on ima nedostatke: računalno je zahtjevниji (faktor  $k$ ), nije zajamčeno da će pronaći globalno najvjerojatniju sekvencu (osim ako  $k$  nije jednak veličini rječnika, što je nepraktično), i još uvjek može generirati tekst koji je relativno generičan, repetitivan ili propušta kreativnije, ali možda malo manje vjerojatne sekvence (Holtzman et al., 2019).

- **Metode uzorkovanja (Sampling Methods):** Kako bi se u generirani tekst unijela veća raznolikost, kreativnost i izbjegla deterministička priroda pohlepnog i beam searcha, koriste se metode koje uvode **slučajnost** u proces odabira sljedećeg tokena, temeljeći se na distribuciji vjerojatnosti koju daje model. Ključne tehnike uključuju:
  - **Temperaturno uzorkovanje (Temperature Sampling):** Prije primjene softmax funkcije za dobivanje konačne distribucije vjerojatnosti, logiti (sirovi izlazi modela prije normalizacije) se dijele s vrijednošću **temperature ( $T$ )**. Ako je  $T = 1$ , distribucija ostaje nepromijenjena. Ako je  $T < 1$  (npr. 0.7), distribucija postaje "oštira", povećavajući vjerojatnosti najvjerojatnijih tokena i smanjujući vjerojatnosti manje vjerojatnih, čineći izlaz sličnjim pohlepnom pretraživanju (konzervativniji, fokusirаниji). Ako je  $T > 1$  (npr. 1.2), distribucija postaje "ravnija", dajući veću šansu i manje vjerojatnim tokenima, što dovodi do raznolikijeg, iznenadujućeg, pa čak i kreativnijeg izlaza, ali uz rizik generiranja nekoherentnog ili besmislenog teksta. Podešavanje temperature omogućuje kontrolu nad ravnotežom između "konzervativnosti" i "kreativnosti" izlaza.
  - **Top-k Uzorkovanje (Top-k Sampling):** Ova metoda (Fan et al., 2018) ograničava uzorkovanje samo na  $k$  najvjerojatnijih tokena prema distribuciji koju daje model. Vjerojatnosti svih ostalih tokena postavljaju se na nulu, a zatim se preostale vjerojatnosti  $k$  tokena renormaliziraju tako da im suma bude 1. Sljedeći token se zatim uzorkuje iz ove reducirane i renormalizirane distribucije. Ovo sprječava odabir vrlo nevjerojatnih ili besmislenih tokena, zadрžavajući pritom određenu razinu slučajnosti i raznolikosti. Međutim, fiksna vrijednost  $k$  može biti problematična: ako je distribucija vrlo "oštra" (model je vrlo siguran u sljedeći token), top-k može i dalje uključivati manje vjerojatne opcije; ako je distribucija vrlo "ravna" (model je nesiguran), top-k može odsjeći mnogo razumnih opcija.
  - **Nucleus (Top-p) Uzorkovanje (Nucleus / Top-p Sampling):** Kako bi se riješio problem fiksne veličine skupa kandidata u top-k, Holtzman et al. (2019) predložili su nucleus sampling. Umjesto odabira fiksнog broja  $k$  tokena, top-p odabire najmanji mogući skup tokena (jezgru, "nucleus") čija **kumulativna vjerojatnost** prelazi unaprijed definirani prag  $p$  (npr.  $p = 0.9$ ). Svi tokeni izvan ove jezgre se odbacuju, a sljedeći token se uzorkuje iz renormalizirane distribucije tokena unutar jezgre. Prednost ovog pristupa je što se veličina skupa kandidata dinamički prilagođava obliku distribucije: kada je model vrlo siguran (oštra distribucija), jezgra će biti mala, uključujući samo nekoliko najvjerojatnijih

tokena; kada je model nesiguran (ravna distribucija), jezgra će biti veća, uključujući više mogućih opcija. Top-p uzorkovanje se pokazalo vrlo efikasnim u generiranju teksta koji je istovremeno koherentan i raznolik, te je postalo vrlo popularna strategija dekodiranja.

U praksi se često koriste kombinacije ovih strategija (npr. top-p uzorkovanje s određenom temperaturom) kako bi se postigao željeni balans između kvalitete, koherentnosti i kreativnosti generiranog teksta.

Konačno, budući da je inferencija često usko grlo u primjeni LLM-ova, posebice u interaktivnim aplikacijama koje zahtijevaju nisku latenciju, značajni napor u ulazu se u **optimizaciju inferencijskog procesa**. Osim već spomenutih tehnika kompresije modela poput kvantizacije i prunninga, ključne optimizacije uključuju:

- **Keširanje Ključeva i Vrijednosti (KV Caching):** Ovo je vjerojatno najvažnija optimizacija za autoregresivno generiranje u Transformerima. Prilikom generiranja svakog novog tokena, model treba izračunati pažnju između trenutnog tokena i svih prethodnih tokena. Izračun Ključ (K) i Vrijednost (V) vektora za prethodne tokene ne ovisi o trenutnom tokenu. Stoga se ovi K i V vektori mogu izračunati samo jednom za svaki prethodni token i zatim **pohraniti u memoriju (keširati)**. Prilikom generiranja sljedećeg tokena, potrebno je izračunati samo Q, K, V za taj novi token i zatim izračunati pažnju koristeći novi Q te nove K, V zajedno s keširanim K i V vektorima iz svih prethodnih koraka. Ovo drastično smanjuje količinu izračuna potrebnih u svakom koraku generiranja, jer se izbjegava ponovno računanje K i V za cijelu prethodnu sekvencu, što dovodi do značajnog ubrzanja inferencije, posebno za duge sekvence.
- **Optimizacije na razini hardvera i softvera:** Korištenje specijaliziranog hardvera (GPU, TPU) s optimiziranim bibliotekama (npr. NVIDIA TensorRT, cuDNN, Google XLA) koje efikasno izvršavaju temeljne operacije poput matričnih množenja. Razvoj specifičnih tehnika poput **FlashAttention** (Dao et al., 2022; Dao, 2023) koje optimiziraju izračun pažnje kako bi se smanjili memorijski prijenosi između GPU memorije, što je često usko grlo.
- **Tehnike paralelizacije:** Za posluživanje vrlo velikih modela koji ne stanu na jedan akcelerator, koriste se različite tehnike paralelizacije, kao što su **tenzorska paralelizacija** (dijeljenje pojedinačnih matričnih operacija preko više uređaja) i **pipeline paralelizacija** (dijeljenje slojeva modela preko više uređaja, pri čemu se podaci obrađuju u stilu pokretne trake) (Shoeybi et al., 2019).
- **Spekulativno Dekodiranje (Speculative Decoding):** Novije tehnike koje pokušavaju ubrzati generiranje korištenjem manjeg, bržeg "nacrt" (draft) modela za brzo generiranje kandidatskih sekvenci od nekoliko tokena, koje zatim veći, sporiji "verifikacijski" model provjerava i prihvaca ili odbacuje u paralelnom koraku (Leviathan et al., 2022; Stern et al., 2018). Ako su predviđanja nacrt modela često točna, ovo može značajno smanjiti broj poziva velikom modelu i ubrzati ukupno vrijeme generiranja.

Dakle, faza inferencije je kritična točka gdje se apstraktne sposobnosti velikih jezičnih modela pretvaraju u konkretnu interakciju i vrijednost. Uspješna inferencija zahtijeva pažljivo balansiranje između kvalitete izlaza, kontrolirane kroz sofisticirane tehnike promptiranja i strategije dekodiranja, i efikasnosti izvršavanja, postignute kroz niz naprednih optimizacijskih

tehnika. Kontinuirani napredak u svim ovim područjima ključan je za daljnje poboljšanje korisničkog iskustva i širenje primjene LLM-ova u sve raznolikijim domenama.

### 3.4 IZAZOVI I OGRANIČENJA: SJENE U DIGITALNOM OGLEDALU

Unatoč impresivnom napretku, LLM-ovi se suočavaju s nizom ozbiljnih izazova i ograničenja koja se ne smiju zanemariti.

#### 3.4.1 Sigurnost, Pouzdanost i Zlouporaba

Dok veliki jezični modeli demonstriraju sve impresivnije sposobnosti u razumijevanju i generiranju jezika, otvarajući vrata transformativnim primjenama u gotovo svim sferama ljudskog djelovanja, njihov brzi razvoj i implementacija neizbjegno su praćeni nizom ozbiljnih izazova i inherentnih ograničenja koja bacaju sjenu na njihov potencijal. Pitanja vezana uz **sigurnost** njihovog korištenja, **pouzdanost** informacija koje generiraju, te potencijal za **zlouporabu** njihovih moćnih sposobnosti, predstavljaju kritične prepreke koje se ne smiju zanemariti. Ovi izazovi nisu tek tehničke poteškoće koje treba prevladati, već duboko isprepletena etička, društvena i sigurnosna pitanja koja zahtijevaju kontinuiranu pažnju, rigorozno istraživanje i odgovorne prakse razvoja i implementacije. Zanemarivanje ovih rizika moglo bi potkopati povjerenje javnosti, dovesti do štetnih posljedica u osjetljivim domenama i ograničiti ostvarivanje punog pozitivnog potencijala ove tehnologije.

Jedan od najpoznatijih i najviše raspravljenih problema vezanih uz pouzdanost LLM-ova jest fenomen **halucinacija**. Termin se odnosi na tendenciju modela da generiraju tekst koji zvuči izuzetno uvjerljivo, koherentno i gramatički ispravno, ali je činjenično netočan, sadrži izmišljene informacije ili čak proturječi samom sebi ili prethodno utvrđenim činjenicama unutar istog odgovora (Ji et al., 2023). Važno je naglasiti da LLM-ovi ne "haluciniraju" u psihološkom smislu; oni nemaju svijest niti uvjerenja koja bi mogla biti pogrešna. Umjesto toga, halucinacije su inherentna posljedica njihove temeljne prirode kao statističkih modela jezika. Oni uče predviđati najvjerojatniji sljedeći token na temelju obrazaca viđenih u ogromnim količinama podataka, ali ne posjeduju eksplizitnu reprezentaciju istinitosti ili činjeničnog znanja odvojenu od tih obrazaca. Halucinacije mogu proizaći iz više razloga: model može reproducirati netočnosti ili pristranosti prisutne u podacima za treniranje, može naučiti lažne korelacije, može "pretjerano samouvjereno" generalizirati izvan područja pokrivenog podacima, ili problemi mogu nastati tijekom samog procesa dekodiranja (npr. strategije uzorkovanja koje favoriziraju fluentnost nad točnošću). Ovaj problem postaje posebno opasan kada se LLM-ovi koriste u domenama gdje je činjenična točnost od presudne važnosti, kao što su medicina (generiranje netočnih medicinskih savjeta ili informacija o lijekovima), pravo (citiranje nepostojećih sudskih presuda ili zakona, kao što je dokumentirano u stvarnim slučajevima gdje su odvjetnici koristili LLM-ove za istraživanje (Schwartz et al., 2023)), financije (davanje pogrešnih finansijskih analiza ili predviđanja) ili obrazovanje (pružanje netočnih povijesnih ili znanstvenih informacija). Borba protiv halucinacija zahtijeva višestruki pristup. Poboljšane tehnike **poravnjanja**, poput RLHF-a ili DPO-a, pokušavaju usaditi modelu sklonost ka "iskrenosti", uključujući sposobnost da prizna neznanje umjesto da izmišlja odgovor. Izuzetno obećavajuća strategija jest **Retrieval-Augmented Generation (RAG)** (Lewis et al., 2020), gdje se proces generiranja odgovora "usidruje" u vanjsku, provjerljivu bazu znanja. Prije generiranja odgovora, RAG sustav prvo pretražuje relevantne dokumente ili podatke iz pouzdanog izvora (npr. interne baze podataka tvrtke, znanstveni članci, provjerene web stranice), a zatim koristi te dohvaćene informacije kao dodatni kontekst za LLM prilikom formuliranja odgovora. Recentna istraživanja u RAG-u fokusiraju se na naprednije tehnike dohvaćanja (npr. iterativno

dohvaćanje, adaptivno dohvaćanje koje odlučuje kada je dohvaćanje uopće potrebno) i bolju integraciju dohvaćenih informacija u proces generiranja (Gao et al., 2023). Nadalje, razvijaju se i mehanizmi za **provjeru činjenica (fact-checking)**, bilo kao post-processing korak ili integrirani u sam model, te prompting tehnike koje potiču model da citira izvore ili eksplisitno izrazi razinu svoje nesigurnosti. Unatoč ovim naporima, potpuno eliminiranje halucinacija ostaje otvoren istraživački problem.

Povezano s halucinacijama, ali s potencijalno izravnijim štetnim namjerama, jest problem **generiranja štetnog sadržaja**. Unatoč značajnim ulaganjima u tehnike poravnanja čiji je cilj učiniti modele bezopasnima, LLM-ovi i dalje mogu biti potaknuti, ili čak spontano generirati, sadržaj koji je toksičan, uvredljiv, mrzilački, diskriminatoran, koji promiče ilegalne aktivnosti, širi dezinformacije ili propagandu. Ovo proizlazi iz činjenice da su modeli trenirani na podacima s interneta koji obiluju takvim sadržajem, te da je definiranje i operacionalizacija pojma "štetnosti" izuzetno složeno i ovisno o kulturnom i društvenom kontekstu. Pokušaji poravnanja da u potpunosti iskorijene ove tendencije često su neuspješni ili nepotpuni, a modeli mogu naučiti samo površinski izbjegavati određene ključne riječi dok zadržavaju sposobnost generiranja štetnog sadržaja na suptilnije načine. Štoviše, postoji stalna "utrka u naoružanju" između developera koji implementiraju sigurnosne mjere i korisnika (ili zlonamjernih aktera) koji pronalaze nove načine kako ih zaobići. Društveni utjecaj generiranja štetnog sadržaja može biti ogroman, uključujući poticanje polarizacije, jačanje štetnih stereotipa, omogućavanje ciljanih kampanja dezinformiranja i pružanje alata za uz nemiravanje ili druge zlonamjerne aktivnosti. Strategije za ublažavanje uključuju rigoroznije **filtriranje podataka** tijekom pre-treniranja, sofisticiranje i **robusnije tehnike poravnanja** (poput Constitutional AI koja pokušava definirati eksplisitne principe (Bai et al., 2022)), implementaciju **višeslojnih filtera sadržaja** koji analiziraju i ulazne promptove i izlazne odgovore modela, te kontinuirano "**red teaming**" – proces gdje stručnjaci aktivno pokušavaju "probiti" sigurnosne mjere modela kako bi identificirali slabosti prije nego što ih zlonamjerni akteri iskoriste (Perez et al., 2022).

Osjetljivost LLM-ova na specifičnu formulaciju ulaznog prompta otvara vrata još jednoj značajnoj sigurnosnoj prijetnji: **zlonamjernim napadima (adversarial attacks)**, posebice kroz tehnike poput **prompt injekcije (prompt injection)** i **jailbreakinga**. Prompt injekcija se odnosi na situaciju gdje napadač uspijeva unutar korisničkog unosa (koji model treba obraditi) umetnuti skrivene instrukcije koje nadjačavaju ili modifiraju originalne upute dane modelu od strane developera (Perez & Ribeiro, 2022; Greshake et al., 2023). U **direktnoj prompt injekciji**, napadač izravno daje maliciozni prompt modelu. U opasnijoj, **indirektnoj prompt injekciji**, maliciozne instrukcije mogu biti skrivene unutar podataka koje model dohvaća iz vanjskih izvora (npr. web stranice, dokumenti) kao dio svog normalnog funkciranja (npr. putem RAG sustava). Model, nesvjestan da obrađuje maliciozni kod unutar naizgled benignog teksta, može biti naveden da izvrši neželjene akcije, poput odavanja povjerljivih informacija iz svoje interne memorije ili konteksta sesije, generiranja štetnog sadržaja, ili čak izvršavanja opasnih naredbi ako je povezan s vanjskim alatima ili API-jima. **Jailbreaking** se odnosi na specifično dizajnirane promptove čiji je cilj eksplisitno zaobići sigurnosne filtere i mjere poravnanja koje su developeri ugradili kako bi spriječili generiranje štetnog sadržaja (Wei et al., 2023; Shen et al., 2023). Ovi promptovi često koriste sofisticirane tehnike, poput uvjerenavljanja modela da igra ulogu lika koji nema etičkih ograničenja, postavljanja problema u hipotetski ili fiktivni kontekst, korištenja enkodinga ili neobičnih formata teksta, ili iskorištavanja inherentne "želje" modela da bude koristan i stijedi upute. Stalno se otkrivaju nove i sve složenije jailbreaking tehnike, čineći obranu izuzetno teškom. Osiguravanje robusnosti modela protiv ovakvih napada ključno je za

njihovu sigurnu primjenu, a strategije obrane uključuju pažljivo **filtriranje i sanitizaciju ulaza, instrukcijsko fino podešavanje (instruction tuning)** kako bi model bolje slijedio sistemske upute i ignorirao konflikte u korisničkom unosu, **adversarijsko treniranje** (izlaganje modela poznatim napadima tijekom treniranja kako bi naučio biti otporniji), te razvoj **specijaliziranih modela ili filtera** dizajniranih za detekciju malicioznih promptova. Unatoč tome, zbog kreativne i neograničene prirode prirodnog jezika, postizanje potpune otpornosti na ove vrste napada ostaje veliki izazov.

Konačno, ispod svih ovih specifičnih problema leži fundamentalno ograničenje: **nedostatak istinskog razumijevanja i sposobnosti rezoniranja** na ljudski način. Unatoč njihovoj sposobnosti da generiraju fluentan i naizgled inteligentan tekst, LLM-ovi su u suštini izuzetno sofisticirani sustavi za prepoznavanje i reprodukciju statističkih obrazaca iz podataka na kojima su trenirani. Oni ne posjeduju svijest, namjere, uvjerenja, niti duboko, **kauzalno razumijevanje svijeta** (Pearl & Mackenzie, 2018; Marcus, 2022). Njihovo "znanje" nije utemeljeno (grounded) u stvarnom iskustvu, percepciji ili interakciji sa svjetom, već proizlazi isključivo iz korelacija i distribucija riječi u tekstualnim podacima (Bender & Koller, 2020 - koncept "stohastičkih papiga"). Ovo inherentno ograničenje manifestira se na više načina: modeli mogu pokazivati nedostatak **zdravorazumskog rezoniranja (common-sense reasoning)**, posebno u situacijama koje nisu eksplicitno ili često videne u podacima za treniranje; mogu imati poteškoća s **kauzalnim zaključivanjem** (razlikovanjem korelacije od uzročnosti); njihova sposobnost **logičke dedukcije** često se temelji na prepoznavanju površinskih obrazaca, a ne na dubokom razumijevanju logičkih pravila; i mogu napraviti absurdne ili **nonsensne pogreške** kada se suoče s ulazima koji su izvan distribucije podataka na kojima su trenirani ili koji sadrže suptilne kontradikcije. Iako napredak u skaliranju i tehnikama poput Chain-of-Thought promptinga poboljšava sposobnosti rezoniranja, ostaje otvoreno pitanje mogu li se istinsko razumijevanje i robusno rezoniranje postići samo dalnjim skaliranjem postojećih arhitektura ili su potrebni fundamentalno novi pristupi. Aktivna područja istraživanja uključuju pokušaje integracije LLM-ova sa **simboličkim metodama rezoniranja ili bazama znanja (knowledge graphs)**, razvoj modela s eksplicitnijim **kauzalnim mehanizmima**, te **utemeljivanje (grounding)** jezičnih modela u **multimodalne podatke** (povezivanje jezika s vidom, slušom, pa čak i akcijom u svijetu) kako bi se stvorila bogatija i kontekstualno utemeljenija reprezentacija značenja.

### 3.4.2 Pristranost, Privatnost i Etika Podataka: Tamna Strana Digitalnih Ogledala

Pored izazova vezanih uz sigurnost i pouzdanost, veliki jezični modeli nasljeđuju i pojačavaju duboko ukorijenjene probleme koji proizlaze iz samih podataka na kojima su stvorenii. Ogromni korupsi teksta i drugih modaliteta, prikupljeni pretežno s interneta, djeluju kao digitalno ogledalo čovječanstva, odražavajući naše znanje i kreativnost, ali i naše predrasude, nejednakosti i etički problematicne prakse. Stoga su pitanja **pristranosti (bias), zaštite privatnosti** i šire **etike korištenja podataka** postala centralna u kritičkom promišljanju razvoja i implementacije LLM-ova. Ova pitanja ne zadiru samo u tehničke fusnote, već u fundamentalno etičko i društveno tkivo pravednosti, jednakosti i ljudskog dostojanstva u doba umjetne inteligencije.

Problem **pristranosti** u velikim jezičnim modelima proizlazi izravno iz činjenice da oni uče iz podataka koje su generirali ljudi, a ti podaci neizbjegno sadrže i reflektiraju postojeće društvene stereotipe, predrasude i sistemske nejednakosti (Blodgett et al., 2020; Bender et al., 2021). Modeli, kao statistički učenici, nisu sposobni inherentno razlikovati poželjne od nepoželjnih obrazaca; oni jednostavno internaliziraju i reproduciraju dominantne korelacije prisutne u

podacima za treniranje. Cijeli taj proces može dovesti do manifestacije različitih vrsta štetnih pristranosti, uključujući **rodne pristranosti** (npr. povezivanje određenih profesija isključivo s jednim spolom, kao što je demonstrirano u radu Bolukbasija i suradnika (2016) na statičkim embeddingsima, a što se prenosi i na LLM-ove), **rasne i etničke pristranosti** (npr. generiranje negativnih stereotipa o određenim skupinama, ili lošije performanse modela na dijalektima ili jezicima manjinskih skupina zbog njihove slabije zastupljenosti u podacima), **kulture** **pristranosti** (npr. favoriziranje zapadnjačkih kulturnih normi i perspektiva), **političke** **pristranosti**, **dobne pristranosti**, i mnoge druge (Mehrabi et al., 2021). Posljedice ovakvih pristranosti mogu biti izuzetno štetne, posebno kada se LLM-ovi koriste u aplikacijama koje utječu na živote ljudi, kao što su sustavi za pomoć pri zapošljavanju (gdje pristran model može favorizirati kandidate određenog spola ili rase), odobravanje kredita, moderiranje sadržaja (gdje može nepravedno cenzurirati glasove manjinskih skupina), ili čak u svakodnevnoj interakciji gdje mogu suptilno, ali kontinuirano pojačavati štetne stereotipe. Mitigacija pristranosti je složen i višeslojan zadatak koji zahtijeva intervencije u različitim fazama životnog ciklusa modela. Počinje s **pažljivom kuracijom i analizom podataka** za treniranje, što uključuje pokušaje identifikacije i filtriranja eksplicitno pristranog sadržaja, ali i sofisticiranje tehnike poput **re-balansiranja podataka** (osiguravanje pravednije zastupljenosti različitih demografskih skupina) ili **augmentacije podataka** (generiranje sintetičkih podataka za podzastupljene skupine). Međutim, definiranje "nepristranog" skupa podataka samo po sebi je normativni izazov. Druga linija obrane uključuje **algoramske tehnike debiasinga**, koje se mogu primijeniti tijekom pre-treniranja (npr. modificiranjem funkcije cilja kako bi se kaznile pristrane korelacije), tijekom finog podešavanja (npr. korištenjem kontrafaktualnih podataka ili tehnika poput adversarijskog debiasinga gdje se model trenira da bude "nesvjestan" osjetljivih atributa (Zhang et al., 2018)), ili čak kao post-processing korak na izlazima modela. Konačno, ključna je **kontinuirana evaluacija pravednosti** koristeći specifične metrike (npr. demografska paritet, jednakost prilika) i specijalizirane benchmarkove dizajnjirane za detekciju stereotipa i pristranosti (npr. StereoSet (Nadeem et al., 2020), CrowS-Pairs (Nangia et al., 2020)). Unatoč ovim naporima, postizanje potpune neutralnosti ili pravednosti ostaje izuzetno teško, dijelom i zbog inherentnih napetosti između različitih definicija pravednosti (Kleinberg et al., 2016).

Drugi kritični etički izazov tiče se **zaštite privatnosti** i fenomena **memorijalizacije podataka**. Veliki jezični modeli, posebice oni trenirani na ogromnim i raznolikim skupovima podataka s interneta, pokazali su zabrinjavajuću sposobnost da nemamjerno "zapamte" (memorijaliziraju) i kasnije, pod određenim uvjetima, doslovno reproduciraju dijelove svojih podataka za treniranje (Carlini et al., 2021). Ovo je posebno problematično kada memorijalizirani podaci uključuju **osjetljive osobne informacije (Personally Identifiable Information - PII)**, kao što su imena, adrese e-pošte, telefonski brojevi, medicinski podaci, finansijske informacije, ili čak jedinstveni dijelovi privatnih razgovora ili tekstova koji su se našli u korpusu za treniranje. Istraživanja su pokazala da modeli mogu biti inducirani da "iscure" ove informacije kroz pažljivo konstruirane promptove (Carlini et al., 2022). Rizik od memorijalizacije veći je za podatke koji su jedinstveni ili se višekratno ponavljaju u skupu za treniranje. Posljedice ovakvog curenja podataka mogu biti ozbiljne, uključujući teške povrede privatnosti pojedinaca, rizik od krađe identiteta, zloupotrebu osjetljivih informacija i kršenje zakona o zaštiti podataka poput GDPR-a. Strategije za ublažavanje rizika memorijalizacije uključuju, prije svega, **rigoroznu sanitizaciju podataka** prije treniranja, s ciljem detekcije i uklanjanja ili anonimizacije PII-a, iako je to izuzetno teško postići sa stopostotnom sigurnošću na tako velikim skupovima podataka. **Deduplikacija podataka** također igra važnu ulogu, jer smanjuje vjerojatnost da model prekomjerno nauči često ponavljane, potencijalno osjetljive sekvence (Lee et al., 2021).

Formalniji pristup zaštiti privatnosti nudi **diferencijalna privatnost (Differential Privacy - DP)** (Dwork et al., 2006). DP je matematički rigorozan okvir koji omogućuje analizu podataka uz pružanje snažnih jamstava da izlaz analize (u ovom slučaju, parametri modela) neće otkriti previše informacija o bilo kojem pojedinačnom zapisu u ulaznim podacima. U kontekstu treniranja dubokih modela, najčešće se koristi tehnika **Diferencijalno Privatnog Stohastičkog Gradijentnog Spusta (DP-SGD)** (Abadi et al., 2016), koja uključuje odsijecanje (clipping) gradjenata pojedinačnih primjera i dodavanje pažljivo kalibriranog Gaussovog šuma prije njihovog agregiranja tijekom procesa treniranja. Iako DP pruža snažna teorijska jamstva privatnosti, njegova primjena na treniranje masivnih LLM-ova suočava se s izazovima: često dolazi uz značajan **pad korisnosti modela** (performanse na nizvodnim zadacima mogu biti znatno lošije), a **računalni troškovi** treniranja s DP-om mogu biti znatno veći. Pronalaženje optimalne ravnoteže između razine privatnosti (kontrolirane parametrom  $\epsilon$ ) i korisnosti modela ostaje ključni istraživački izazov (Li et al., 2021). Konačno, važnu ulogu igraju i **kontrole pristupa i sigurne prakse implementacije** modela kako bi se spriječio neovlašteni pristup modelu ili njegovim izlazima koji bi mogli sadržavati memorijalizirane podatke.

Nadalje, inherentna složenost i veličina velikih jezičnih modela dovode do značajnih izazova u pogledu **transparentnosti i objašnjivosti (Explainability - XAI)**. LLM-ovi se često opisuju kao "**crne kutije**" – njihove unutarnje operacije, koje uključuju milijarde ili trilijune parametara organiziranih u duboke i složene neuronske arhitekture, izuzetno su teške za ljudsku interpretaciju. Teško je, ako ne i nemoguće, precizno razumjeti **zašto** je model generirao određeni odgovor, na temelju kojih "znanja" ili obrazaca je donio odluku, ili kako bi se njegovo ponašanje moglo promijeniti s malim varijacijama u ulazu. Ovaj nedostatak transparentnosti predstavlja ozbiljnu prepreku iz više razloga: otežava **debugiranje** modela i identifikaciju uzroka pogrešaka ili neželjenih ponašanja (poput halucinacija ili pristranosti); potkopava **povjerenje korisnika**, posebno u visoko rizičnim primjenama; otežava osiguravanje **odgovornosti** za odluke donesene uz pomoć AI sustava; i može biti u sukobu s regulatornim zahtjevima, poput "prava na objašnjenje" koje impliciraju neki zakoni o zaštiti podataka (npr. GDPR). Razvoj metoda za objašnjivost specifično za LLM-ove je aktivno, ali još uvjek relativno nezrelo područje istraživanja (Danilevsky et al., 2020). Pristupi uključuju **vizualizaciju mehanizama pažnje** (iako je interpretacija težina pažnje kao direktnog objašnjenja često kritizirana kao potencijalno zavaravajuća (Jain & Wallace, 2019; Wiegrefe & Pinter, 2019)), korištenje **tehnika atribucije značajki (feature attribution)** poput LIME-a ili SHAP-a prilagođenih za tekst kako bi se identificirali najutjecajniji ulazni tokeni za određeni izlaz, **analizu internih reprezentacija** modela (npr. treniranjem jednostavnih "sondi" (probing classifiers) za ispitivanje koje su lingvističke ili semantičke informacije kodirane u različitim slojevima (Belinkov & Glass, 2019)), te novije tehnike usmjerene na **kauzalno praćenje i uređivanje modela (causal tracing and model editing)** koje pokušavaju locirati i čak modificirati specifična "znanja" ili ponašanja unutar mreže (Meng et al., 2022). Neki pristupi također pokušavaju istrenirati model da sam **generira objašnjenja** u prirodnom jeziku za svoje odgovore, često u kombinaciji s tehnikama poput Chain-of-Thought. Unatoč ovim naporima, pružanje vjernih, razumljivih i korisnih objašnjenja za kompleksne odluke LLM-ova ostaje fundamentalni izazov.

Konačno, sama osnova postojanja velikih jezičnih modela – njihovo treniranje na ogromnim količinama podataka prikupljenih s interneta – otvara složena i kontroverzna pravna i etička pitanja vezana uz **autorska prava (copyright)** i **vlasništvo nad podacima**. Velik dio tekstualnih, slikovnih i kodnih podataka korištenih za pre-treniranje zaštićen je autorskim pravima. Praksa "struganja" (scraping) ovih podataka s weba bez eksplicitne dozvole ili naknade nositeljima

prava dovela je do niza **pravnih sporova** visokog profila, gdje autori, umjetnici, novinske kuće (npr. The New York Times vs. OpenAI & Microsoft) i drugi kreatori sadržaja tuže AI kompanije tvrdeći da treniranje modela na njihovim djelima predstavlja kršenje autorskih prava (Sanderson, 2023; Getty Images vs. Stability AI). Ključno pravno pitanje, posebice u američkom pravnom sustavu, jest može li se takvo korištenje smatrati "**pravednom upotreborom**" (**fair use**), doktrinom koja dopušta ograničeno korištenje zaštićenog materijala bez dozvole za svrhe poput kritike, komentara, izvještavanja, podučavanja ili istraživanja. AI kompanije često argumentiraju da je treniranje transformativna upotreba koja stvara nešto novo i ne zamjenjuje izravno originalna djela, te da je neophodno za tehnološki napredak. Kreatori, s druge strane, ističu da modeli uče iz njihovog rada bez kompenzacije i mogu generirati izlaze koji se izravno natječu s njihovim djelima ili čak reproduciraju njihov stil. Ishodi ovih sudskih sporova i razvoj zakonodavstva u ovom području (poput EU AI Acta koji uvodi određene obveze transparentnosti u pogledu podataka za treniranje) imat će dubok utjecaj na budućnost razvoja LLM-ova. Potencijalna rješenja koja se istražuju uključuju razvoj **sustava licenciranja** za podatke za treniranje, uspostavu **mehanizama za isključivanje (opt-out)** koji bi omogućili kreatorima da sprječe korištenje svojih djela za treniranje, te razvoj modela treniranih isključivo na **provjereni licenciranim ili javno dostupnim podacima**, iako to može ograničiti njihove sposobnosti ili povećati troškove. Pitanje pravedne kompenzacije za podatke koji pokreću AI revoluciju ostaje jedno od najspornijih etičkih i ekonomskih pitanja današnjice.

### 3.4.3 Resursna Intenzivnost: Cijena Inteligencije

Dok nas zadivljuju svojim jezičnim sposobnostima, veliki jezični modeli dolaze s enormnom, često prešućivanom cijenom. Njihova impresivna inteligencija i performanse izravno su povezane s njihovom monumentalnom veličinom i složenošću, što ih čini izuzetno **resursno intenzivnim** sustavima. Ova intenzivnost manifestira se kroz tri ključna aspekta: astronomske **računalne i financijske troškove** potrebne za njihovo treniranje, zabrinjavajući **energetski otisak i okolišni utjecaj**, te ogromne **memorijske zahtjeve za** njihovo pokretanje. Ovi faktori ne predstavljaju samo tehničke izazove za inženjere, već imaju duboke ekonomске, društvene i etičke implikacije, oblikujući tko može razvijati i koristiti ovu moćnu tehnologiju, te postavljajući ozbiljna pitanja o dugoročnoj održivosti trenutne putanje razvoja umjetne inteligencije.

Prvi i najočitiji trošak povezan je s **treniranjem** najnaprednijih LLM-ova. Proces pre-treniranja, tijekom kojeg model uči temeljne obrasce jezika iz petabajta podataka, zahtijeva nezamislive količine računalne snage. Izvođenje milijardi milijardi matematičkih operacija (FLOPs - Floating Point Operations Per Second) potrebnih za optimizaciju trilijuna parametara kroz više epoha treniranja moguće je samo korištenjem masivnih klastera specijaliziranog hardvera, primarno grafičkih procesorskih jedinica (GPU) visokih performansi (poput NVIDIA-inih A100 ili H100 čipova) ili tenzorskih procesorskih jedinica (TPU) koje razvija Google (Patterson et al., 2021). Ovi klasteri mogu sadržavati tisuće ili čak desetke tisuća pojedinačnih akceleratora koji rade paralelno tjednima ili mjesecima bez prestanka. Nabava, održavanje i pogon takve infrastrukture rezultira izravnim **financijskim troškovima** koji se mijere u **desecima ili čak stotinama milijuna dolara** za treniranje samo jednog state-of-the-art modela (procjene za modele poput GPT-4 idu i preko 100 milijuna dolara, iako točne brojke rijetko bivaju objavljene). Ovako visoka cijena stvara ogromnu barjeru ulasku na tržište. Samo nekolicina najvećih svjetskih tehnoloških kompanija (poput Googlea, Microsoft-a/OpenAI-ja, Mete, Anthropic-a) posjeduje financijske i infrastrukturne resurse potrebne za razvoj i treniranje modela na samom

vrhu ljestvice performansi. Ovo neizbjegno dovodi do **koncentracije moći** i kontrole nad razvojem najnaprednije AI tehnologije u rukama malog broja aktera, potencijalno gušeći konkureniju i inovacije koje bi mogle doći iz akademske zajednice, manjih tvrtki ili neprofitnih organizacija, koje si jednostavno ne mogu priuštiti takva ulaganja. Stvara se "računalna podjela" (compute divide) koja može produbiti postojeće nejednakosti u pristupu tehnologiji i njenim beneficijama.

Drugi, sve više prepoznat problem jest ogroman **energetski otisak i povezani okolišni utjecaj** velikih jezičnih modela. Intenzivno i dugotrajno korištenje tisuća energetski gladih GPU/TPU čipova tijekom treniranja troši ogromne količine električne energije. Studije koje su pokušale procijeniti ovaj utjecaj došle su do zabrinjavajućih brojki. Na primjer, Strubell i suradnici (2019) procijenili su da treniranje jednog tadašnjeg velikog Transformer modela (s otprilike 213 milijuna parametara, znatno manje od današnjih modela) može emitirati količinu ugljičnog dioksida usporedivu s pet prosječnih automobila tijekom njihovog cijelog životnog vijeka, uključujući proizvodnju. Novije studije, poput one Luccionija i suradnika (2022) koja je analizirala model BLOOM (176 miljardi parametara), procijenile su emisije CO<sub>2</sub> tijekom njegovog treniranja na oko 25 metričkih tona ekvivalenta CO<sub>2</sub>, što je i dalje značajna brojka. Iako je važno napomenuti da ove procjene ovise o mnogim faktorima (poput specifičnog hardvera, efikasnosti softvera, trajanja treniranja i, ključno, izvora električne energije korištene u podatkovnim centrima – je li iz fosilnih goriva ili obnovljivih izvora), one jasno ukazuju da razvoj LLM-ova ima mjerljiv doprinos emisijama stakleničkih plinova i klimatskim promjenama. Nadalje, dok se fokus često stavlja na jednokratne, ali masivne troškove treniranja, ne smije se zanemariti ni kumulativni energetski trošak **inferencije**. Iako je generiranje odgovora za pojedinačni upit energetski znatno manje zahtjevno od treniranja, široka i kontinuirana upotreba LLM-ova od strane milijuna ili miljardi korisnika diljem svijeta znači da ukupna potrošnja energije za inferenciju može s vremenom premašiti onu utrošenu na treniranje (Luccioni et al., 2023). Rastuća svijest o klimatskoj krizi stavlja sve veći pritisak na AI zajednicu i industriju da razvijaju **održive prakse**. To uključuje ne samo napore u **optimizaciji efikasnosti** samih modela i algoritama (kako je opisano u prethodnom odjeljku), već i transparentnije izvještavanje o potrošnji energije i emisijama, te, što je najvažnije, korištenje **obnovljivih izvora energije** za napajanje podatkovnih centara gdje se modeli treniraju i pokreću.

Treći aspekt resursne intenzivnosti odnosi se na ogromne **memorijske zahtjeve** velikih jezičnih modela. Ovi zahtjevi proizlaze iz dva glavna izvora. Prvi je potreba za pohranom samih **parametara modela**. Model sa stotinama milijardi ili trilijunima parametara zahtjeva stotine ili tisuće gigabajta memorije samo za smještaj tih težina (npr., model sa 175 milijardi parametara, ako se svaki parametar pohranjuje s 16-bitnom preciznošću (FP16), zahtjeva oko 350 GB memorije). Drugi izvor su **privremene aktivacije** koje se moraju čuvati u memoriji tijekom procesa izračuna. Tijekom faze treniranja (posebno kod korištenja backpropagation algoritma) i tijekom autoregresivnog generiranja u fazi inferencije, potrebno je pohraniti međurezultate iz različitih slojeva modela. Posebno kod inferencije dugih sekvenci, **KV keš (Key-Value Cache)**, koji pohranjuje izračunate Ključ i Vrijednost vektore iz mehanizma pažnje za prethodne tokene kako bi se ubrzalo generiranje sljedećih, može narasti do vrlo velikih dimenzija, često premašujući memoriju potrebnu za same parametre modela (Pope et al., 2023). Ovi visoki memorijski zahtjevi, posebice za specijaliziranu memoriju visokih performansi na GPU/TPU akceleratorima (VRAM ili HBM - High Bandwidth Memory), koja je znatno skupljia i kapacitetom ograničenija od standardne sistemske RAM memorije, predstavljaju značajno usko grlo. Posljedica je da se najveći modeli često ne mogu pokrenuti na jednom akceleratoru,

već zahtijevaju distribuciju preko više njih koristeći složene tehnike paralelizacije. Što je još važnije za širu primjenu, ovi memoriski zahtjevi čine izuzetno teškim, ako ne i nemogućim, pokretanje najnaprednijih LLM-ova izravno na **uređajima s ograničenim resursima**, kao što su pametni telefoni, prijenosna računala bez dediciranih snažnih GPU-ova, ili rubni (edge) uređaji u IoT okruženjima. Ovo ograničava dostupnost naprednih AI mogućnosti "offline" i u scenarijima gdje je niska latencija ključna, a oslanjanje na cloud infrastrukturu nije uvijek poželjno ili moguće zbog privatnosti, cijene ili povezanosti. Tehnike optimizacije poput kvantizacije (koja izravno smanjuje memoriju potrebnu za parametre) i pruninga, kao i razvoj manjih, ali i dalje sposobnih modela (npr. tzv. SLM - Small Language Models), ključne su za ublažavanje ovog memoriskog ograničenja i omogućavanje šire primjene LLM tehnologije.

Suočavanje s ovim višestrukim izazovima resursne intenzivnosti – visokim financijskim troškovima, značajnim energetskim otiskom i ogromnim memoriskim zahtjevima – nije jednostavan zadatak i zahtjeva **multidisciplinarni pristup**. Potrebne su daljnje **tehničke inovacije** u razvoju efikasnijih algoritama za treniranje i inferenciju, efikasnijih arhitektura modela, te energetski efikasnijeg hardvera. Istovremeno, potrebne su jasne **etičke smjernice** i potencijalno **regulatorni okviri** koji bi poticali transparentnost u izvještavanju o resursnoj potrošnji i okolišnom utjecaju, te promovirali razvoj održivijih AI praksi. Neophodan je i **kontinuirani dijalog** između istraživača, developera, kompanija, donositelja politika i šire javnosti kako bi se postigao konsenzus o prihvatljivoj cijeni umjetne inteligencije i osiguralo da koristi ove moćne tehnologije budu dostupne što širem krugu ljudi na pravedan i održiv način. Samo kroz zajedničke napore možemo osigurati da "cijena inteligencije" ne postane previsoka za naše društvo i naš planet.

### 3.5 TRANSFORMACIJA KOMUNIKACIJE: PRIMJENE I PERSPEKTIVE

Unatoč inherentnim izazovima i ograničenjima detaljno opisanim ranije, utjecaj velikih jezičnih modela na tkivo ljudske komunikacije, radne procese i pristup informacijama već je sada dubok, sveprisutan i ubrzava se eksponencijalnom brzinom. Daleko od toga da su samo laboratorijski kurioziteti, LLM-ovi su postali pokretačka snaga iza nove generacije aplikacija koje redefiniraju interakciju čovjeka i stroja te otvaraju neslućene mogućnosti u gotovo svakom zamislivom području. Njihove primjene nisu samo raznolike, već se i neprestano šire i produbljuju, potaknute kontinuiranim napretkom u arhitekturama modela, tehnikama treniranja i, što je ključno, rastućom dostupnošću i integracijom u postojeće platforme i alate. Ova transformacija nije samo inkrementalna; ona predstavlja fundamentalni pomak u načinu na koji stvaramo, dijelimo, konzumiramo i razumijemo informacije i međusobno komuniciramo.

Jedno od najočitijih i najbrže rastućih područja primjene jest **razgovorni AI (Conversational AI)**. Moderni LLM-ovi, poput OpenAI-jevog GPT-4o (OpenAI, 2024b), Anthropicovog Claude 3 obitelji (Anthropic, 2024), Googleovog Gemini 1.5 Pro (Google DeepMind, 2024) i Metinog Llama 3 (Meta AI, 2024), pokreću chatbotove i virtualne asistente koji su svjetlosnim godinama ispred svojih prethodnika temeljenih na pravilima ili jednostavnijim AI tehnikama. Ovi sustavi sposobni su voditi iznenađujuće prirodne, koherentne i kontekstualno svjesne razgovore o širokom spektru tema. Mogu odgovarati na složena pitanja koja zahtijevaju sintezu informacija, pomagati korisnicima u kreativnim i tehničkim zadacima poput pisanja eseja, generiranja programskog koda ili učenja novih vještina. Najnoviji modeli pokazuju i poboljšano razumijevanje nijansi, humora i čak implicitnih namjera u razgovoru, približavajući se fluidnosti ljudske interakcije. U poslovnom svijetu, LLM-pokretani chatbotovi transformiraju korisničku podršku, pružajući trenutne, personalizirane odgovore 24/7 i rješavajući sve složenije upite,

čime oslobođaju ljudske agente za zahtjevније probleme (Accenture, 2023). Virtualni asistenti integrirani u operativne sustave i pametne uređaje postaju proaktivniji, sposobni ne samo reagirati na naredbe, već i predviđati potrebe korisnika i nuditi relevantne informacije ili pomoći. Štoviše, svjedočimo usponu **specijaliziranih AI agenata** temeljenih na LLM-ovima, dizajniranih za specifične komunikacijske zadatke, poput pregovaranja, prodaje, ili čak pružanja podrške u području mentalnog zdravlja (iako ovo posljednje otvara ozbiljna etička pitanja i zahtjeva izuzetan oprez) (D'Alfonso, 2023).

Usko povezano s razgovornim sposobnostima jest područje **generiranja i sažimanja sadržaja**. LLM-ovi djeluju kao moćni alati za automatizaciju i augmentaciju procesa stvaranja teksta. U novinarstvu i marketingu koriste se za generiranje nacrta članaka, priopćenja za javnost, marketinških kopija, opisa proizvoda ili postova za društvene medije, značajno ubrzavajući produkciju sadržaja (Salesforce, 2024). Alati poput Jaspera, Copy.ai ili integriranih AI funkcija u platformama kao što su Microsoft 365 Copilot ili Google Workspace Duet AI djeluju kao "kopiloti" za pisanje, nudeći sugestije, preoblikujući tekst ili generirajući cijele odlomke na temelju kratkih uputa. Sposobnost LLM-ova da **sažimaju** duge dokumente, transkripte sastanaka, lance e-pošte ili znanstvene radove u kratke, informativne pregledе štedi vrijeme i olakšava snalaženje u poplavi informacija. U kreativnim industrijama, pisci i scenaristi koriste LLM-ove kao partnere za brainstorming, generiranje ideja, razvoj likova ili prevladavanje spisateljske blokade. Međutim, oslanjanje na AI za generiranje sadržaja također postavlja pitanja o originalnosti, autentičnosti, potencijalnom širenju dezinformacija (ako model halucinira) i potrebi za pažljivim ljudskim uređivanjem i provjerom činjenica. Kvaliteta AI-generiranog sadržaja može varirati, a prekomjerna upotreba može dovesti do homogenizacije stila i gubitka jedinstvenog ljudskog glasa.

Područje **strojnog prevodenja (Machine Translation - MT)** doživjelo je kvantni skok s pojmom neuroninskih mreža (Neural Machine Translation - NMT), a LLM-ovi su dodatno podigli ljestvicu kvalitete, fluentnosti i kontekstualne prikladnosti prijevoda (Koehn, 2020; Brown et al., 2020). Modeli trenirani na masivnim višejezičnim korpusima pokazuju znatno bolje razumijevanje idioma, kulturnih nijansi i složenih gramatičkih struktura u usporedbi s ranijim statističkim ili NMT sustavima. Usluge poput Google Prevoditelja, DeepL-a i Microsoft Prevoditelja intenzivno koriste LLM tehnologije kako bi pružile prijevode koji su često teško razlikovni od ljudskih za mnoge jezične parove. Posebno uzbudljiv razvoj je **prevodenje u stvarnom vremenu**, omogućeno napretkom u prepoznavanju govora, LLM prevodenju i sintezi govora. Najnoviji multimodalni modeli poput GPT-4o demonstriraju sposobnost gotovo trenutnog prevodenja govornog jezika, uz očuvanje tona i emocija govornika, otvarajući vrata besprijeckornoj komunikaciji na sastancima, putovanjima ili u svakodnevnim interakcijama (OpenAI, 2024a). Značajni napor u ulazu se i u poboljšanje prevodenja za **jezike s malo resursa (low-resource languages)**, za koje tradicionalno nedostaje velikih paralelnih korpusa. Tehnike poput transfer learninga, višejezičnog pre-treniranja (npr. Meta-in projekt NLLB - No Language Left Behind (Costa-jussà et al., 2022)) i korištenja LLM-ova za generiranje sintetičkih podataka pomažu u premošćivanju ovog jaza, čineći blagodati prevodenja dostupnijima širem krugu govornika.

LLM-ovi su također postali nezamjenjivi alati za **analizu jezika** na velikoj skali. Njihova sposobnost razumijevanja semantike i konteksta omogućuje sofisticiranu **analizu sentimenta** koja ide dalje od jednostavne pozitivno/negativno klasifikacije, prepoznajući nijansiranje emocije, sarkazam ili ironiju u recenzijama proizvoda, objavama na društvenim mrežama ili odgovorima na ankete. Ovo pruža tvrtkama dublji uvid u percepciju brenda i zadovoljstvo korisnika (The Economist, 2023). **Prepoznavanje imenovanih entiteta**

(NER) i ekstrakcija odnosa (relation extraction) omogućuju automatsko izvlačenje strukturiranih informacija iz nestrukturiranog teksta, što je ključno u područjima poput poslovne inteligencije (analiza vijesti o konkurenčiji), znanstvenih istraživanja (ekstrakcija podataka iz literature) i legal tech-a (analiza ugovora, e-discovery – pretraživanje i klasifikacija ogromnih količina pravnih dokumenata) (Tallinn University of Technology, 2024).

U kibernetičkoj sigurnosti, LLM-ovi se koriste za analizu logova sustava, izvještaja o prijetnjama i komunikacije na dark webu kako bi se identificirali potencijalni napadi ili sigurnosni propusti. **Moderiranje sadržaja** na online platformama također se sve više oslanja na LLM-ove za detekciju govora mržnje, uznemiravanja, dezinformacija i drugog štetnog sadržaja, iako je postizanje pravedne i konzistentne moderacije i dalje veliki izazov zbog kontekstualne osjetljivosti i potencijalnih pristranosti modela.

Sposobnost LLM-ova da razumiju individualne preferencije i kontekst korisnika pokreće novu eru **hiper-personalizacije**. U marketingu i e-trgovini, to znači kreiranje dinamički prilagođenih preporuka proizvoda, marketinških poruka, pa čak i izgleda web stranica temeljenih na povijesti pregledavanja, kupovine i drugim signalima korisnika (McKinsey, 2023). Medijske platforme koriste LLM-ove za kuriranje personaliziranih vijesti i sadržaja. U **obrazovanju**, ovo omogućuje razvoj **adaptivnih platformi za učenje** koje prilagođavaju tempo, stil i sadržaj podučavanja individualnim potrebama i sposobnostima svakog učenika (UNESCO, 2023). U **zdravstvu**, personalizacija se može odnositi na prilagodene komunikacijske strategije za pacijente ili podsjetnike za uzimanje lijekova. Međutim, ova moć personalizacije dolazi s etičkim odgovornostima. Postoji tanka linija između korisne personalizacije i manipulativnog ciljanja ili stvaranja **filter mješurića (filter bubbles)** koji ograničavaju izloženost korisnika različitim perspektivama. Transparentnost i kontrola korisnika nad vlastitim podacima ključni su za odgovornu primjenu.

Utjecaj LLM-ova na **obrazovanje** je višestruk i predmet intenzivne rasprave. S jedne strane, oni nude ogroman potencijal kao **personalizirani AI tutori** koji mogu pružiti individualiziranu pomoć učenicima, odgovarati na pitanja, objašnjavati koncepte i pružati vježbe prilagođene njihovom tempu (Khan Academy koristi GPT-4 za svog Khanmigo tutora (Khan Academy, 2023)). Mogu pomoći nastavnicima u **automatizaciji administrativnih zadataka**, poput generiranja nacrta nastavnih planova ili čak pružanja pomoći pri ocjenjivanju eseja (iako s oprezom). Učenicima služe kao moćni alati za **istraživanje, pisanje** (pomoći pri strukturiranju, preoblikovanju, provjeri gramatike) i **učenje jezika**. S druge strane, postoje značajne zabrinutosti oko **prekomjernog oslanjanja** učenika na AI alate, potencijala za **akademsko nepoštenje (plagijarizam, varanje)**, **produbljivanja digitalnog jaza** (ako pristup alatima nije jednak za sve) i potrebe za razvojem **kritičke AI pismenosti** kod učenika i nastavnika kako bi mogli odgovorno koristiti ove tehnologije (Miao et al., 2024; UNESCO, 2023). Integracija LLM-ova u obrazovanje zahtijeva pažljivo promišljanje pedagoških pristupa i etičkih smjernica.

U **zdravstvu**, LLM-ovi pokazuju obećavajuće rezultate u nizu primjena, iako je ovdje potreban najveći oprez zbog potencijalnih posljedica pogrešaka. Koriste se za **analizu ogromnih količina medicinske literature** kako bi se ubrzala istraživanja i informirale kliničke odluke. Mogu pomoći lijećnicima u **sažimanju elektroničkih zdravstvenih kartona (EHR)** ili generiranju nacrta kliničkih bilješki, potencijalno smanjujući administrativno opterećenje i **liječničko izgaranje (burnout)** (Microsoft + Nuance, 2023). Postoje istraživanja koja pokazuju potencijal LLM-ova u podršci **dijagnostičkim procesima**, na primjer, analizom simptoma opisanih od strane pacijenta ili interpretacijom nalaza (iako ne kao zamjena za liječničku prosudbu) (Thirunavukarasu et al., 2023). U **farmaceutskoj industriji**, koriste se za ubrzavanje **otkrivanja**

**lijekova** analizom podataka o molekulama i biološkim putevima. Također se razvijaju alati za **komunikaciju s pacijentima**, pružanje zdravstvenih informacija na razumljiv način ili odgovaranje na česta pitanja. Međutim, apsolutna nužnost **činjenične točnosti, zaštite privatnosti pacijenata (HIPAA i sl.), transparentnosti, robusnosti i kontinuirane validacije** u kliničkim uvjetima, uz neizostavan **ljudski nadzor**, ključni su preduvjeti za sigurnu i etičku primjenu LLM-ova u zdravstvu (Nature Medicine Editorial, 2023).

Sektori **prava i financija** također doživljavaju značajne promjene. U pravu, LLM-ovi se koriste za **automatizaciju pregleda i analize ugovora**, identifikaciju ključnih klauzula ili potencijalnih rizika, **ubrzavanje pravnog istraživanja** pretraživanjem i sažimanjem relevantne sudske prakse i zakona, **pomoć pri izradi nacrta pravnih dokumenata** (npr. podnesaka, dopisa), pa čak i za **prediktivnu analitiku** ishoda sporova (iako s ograničenom pouzdanošću). U financijama, LLM-ovi analiziraju finansijske vijesti, izvještaje kompanija i sentiment na društvenim mrežama kako bi generirali **signale za algoritamsko trgovanje**, automatiziraju **analizu finansijskih izvještaja**, pomažu u **detekciji prijevara** analizom transakcijskih obrazaca i komunikacije, te pokreću chatbotove koji pružaju **personalizirane finansijske savjete** ili korisničku podršku (Deloitte, 2023). Kao i u zdravstvu, regulatorni zahtjevi, potreba za točnošću i etička razmatranja (npr. pravednost u davanju savjeta) ključni su u ovim visoko reguliranim industrijama.

Razvoj **softvera** jedno je od područja gdje je utjecaj LLM-ova možda najizravniji i najbrže prihvaćen. Alati poput **GitHub Copilota** (koji koristi OpenAI Codex, izvedenicu GPT modela), Amazonovog **CodeWhisperera**, Googleovog **Duet AI for Developers** i drugih djeluju kao **AI asistenti za kodiranje**, pružajući programerima sugestije za dovršavanje koda u stvarnom vremenu, generirajući cijele funkcije ili blokove koda na temelju komentara u prirodnom jeziku, pomažući u **debugiranju** pronalaženjem i predlaganjem ispravaka za pogreške, **automatizirajući pisanje testova**, pa čak i **prevodeći kod** između različitih programskih jezika (Chen et al., 2021; Ziegler et al., 2022). Ovi alati značajno povećavaju **produktivnost programera**, smanjuju repetitivne zadatke i mogu pomoći u učenju novih jezika ili okvira. Također omogućuju nove načine interakcije s kodom, poput postavljanja pitanja o funkcionalnosti koda u prirodnom jeziku. Iako ne zamjenjuju potrebu za ljudskim razumijevanjem i arhitektonskim odlukama, oni mijenjaju prirodu softverskog inženjerstva, čineći ga više suradničkim procesom između čovjeka i AI-ja.

U domeni **znanstvenih istraživanja**, LLM-ovi postaju moćni saveznici znanstvenicima. Koriste se za **ubrzavanje pregleda literature** sažimanjem velikog broja članaka ili identificiranjem relevantnih radova. Mogu pomoći u **analizi velikih skupova nestrukturiranih podataka**, poput tekstualnih odgovora u anketama, terenskih bilješki ili čak podataka iz eksperimentenata (npr. analiza sekvenci u genomici). Neki istraživači eksperimentiraju s korištenjem LLM-ova za **generiranje hipoteza** na temelju postojeće literature ili za **predlaganje dizajna eksperimenta**. Također pružaju pomoć pri **pisanju znanstvenih radova**, od generiranja nacrta određenih sekcija do poboljšanja jasnoće i gramatike. Platforme poput Elicit.org ili Scite.ai koriste LLM-ove kako bi pomogle istraživačima u pronalaženju informacija i razumijevanju znanstvene literature na nove načine. Naravno, i ovdje je ključna ljudska ekspertiza za kritičku procjenu, validaciju i osiguravanje znanstvene rigoroznosti.

Posebno značajan i recentan trend jest uspon **multimodalnih LLM-ova**, koji mogu obrađivati i generirati informacije u više modaliteta istovremeno – ne samo tekst, već i **slike, zvuk i video**. Modeli poput Googleovog Gemini 1.5 Pro (Google DeepMind, 2024) i OpenAI-jevog GPT-4o (OpenAI, 2024b) demonstriraju impresivne sposobnosti u ovom području. Mogu voditi **govorni razgovor u stvarnom vremenu** dok istovremeno **analiziraju vizualni unos** s kamere (npr.

opisujući korisniku okolinu, čitajući jelovnik, pomažući u rješavanju matematičkog problema napisanog na papiru). Mogu generirati slike iz detaljnih tekstualnih opisa (nadvozujući se na rad modela poput DALL-E ili Midjourney), ali i opisivati slike ili odgovarati na pitanja o njima s velikom preciznošću. Mogu analizirati video sadržaj, transkribirati govor iz audio zapisa ili čak generirati glazbu. Ova sposobnost integracije različitih modaliteta otvara vrata potpuno novim oblicima prirodnije i intuitivnije interakcije čovjeka i računala, te omogućuje primjene koje su prije bile nezamislive, poput naprednih alata za osobe s invaliditetom (npr. opisivanje vizualnog svijeta slijepim osobama), sofisticiranje analize medicinskih snimki u kombinaciji s tekstualnim izvještajima, ili stvaranja bogatijih i interaktivnijih edukativnih i zabavnih sadržaja. Multimodalnost predstavlja ključni korak prema AI sustavima koji mogu percipirati i komunicirati o svijetu na način bliži ljudskom iskustvu.

Svi ovi napretci pokretani su kontinuiranim razvojem vodećih modela, pri čemu kompanije poput OpenAI-ja (GPT-4, GPT-4o), Anthropic-a (Claude 3 obitelj - Haiku, Sonnet, Opus), Googlea (Gemini obitelj - Nano, Flash, Pro, Ultra), Mete (Llama 3 serija, uključujući i modele otvorenog koda) i Mistral AI-ja (Mistral Large i modeli otvorenog koda) neprestano pomicu granice mogućeg u pogledu performansi, efikasnosti i multimodalnih sposobnosti. Natjecanje, ali i sve veća dostupnost moćnih modela (posebice onih otvorenog koda poput Llama 3 ili Mistral modela), ubrzavaju inovacije i širenje primjena.

Gledajući u budućnost, jasno je da transformacija komunikacije potaknuta LLM-ovima tek uzima maha. Sljedeći logičan korak, koji je već u tijeku, jest razvoj autonomnijih AI agenata. Ovi agenti koriste LLM-ove kao svoj "mozak" ili kontrolni mehanizam, dajući im sposobnost ne samo da razumiju i generiraju jezik, već i da planiraju, rezoniraju i poduzimaju akcije u digitalnom ili čak fizičkom svijetu kako bi postigli zadane ciljeve (Wang et al., 2023; Xi et al., 2023). Takvi agenti bi mogli samostalno obavljati složene zadatke poput organizacije putovanja, upravljanja kalendarom, provođenja online istraživanja, interakcije s drugim softverskim alatima ili čak upravljanja pametnim uređajima. Upravo će uspon ovih AI agenata, koji se oslanjaju na komunikacijske i kognitivne sposobnosti LLM-ova, predstavljati sljedeću fazu revolucije i biti centralna tema dalnjih razmatranja u ovoj knjizi. Naravno, ovaj razvoj donosi i nove, još veće izazove vezane uz kontrolu, sigurnost, etiku i potencijalni utjecaj na društvo i zapošljavanje, koji zahtijevaju proaktivno promišljanje i usmjeravanje.

### 3.6 LLM: SLJEDEĆA STANICA EVOLUCIJE?

Veliki jezični modeli predstavljaju tehnološki skok kvantnog reda veličine u sposobnosti strojeva da barataju ljudskim jezikom. Prošli smo put od jednostavnih statističkih modela do složenih Transformer arhitektura koje uče iz digitalnih prostranstava podataka. Razumjeli smo njihov životni ciklus – od intenzivnog pre-treniranja, preko ključnog finog podešavanja i poravnanja s ljudskim vrijednostima, do optimizacija koje ih čine efikasnijima i evaluacija koje pokušavaju izmjeriti njihove stvarne sposobnosti.

Međutim, suočili smo se i s tamnjom stranom: izazovima sigurnosti, pouzdanosti, pristranosti, privatnosti i ogromnim troškovima. Ovi izazovi naglašavaju da LLM-ovi nisu čarobni štapići, već moći alati čiji razvoj i primjenu moramo voditi mudro i odgovorno.

Njihov utjecaj na komunikaciju je neosporan i tek počinje. Automatizacija analize i generiranja jezika, prevođenje u stvarnom vremenu i nove mogućnosti multimodalne interakcije transformiraju profesije, industrije i našu svakodnevnu interakciju s tehnologijom.

LLM-ovi nisu kraj komunikacijske evolucije, ali jesu njezina važna, možda i prijelomna, sljedeća stanica. Oni su temelj za razvoj još naprednijih sustava – **AI agenata** – autonomnih entiteta koji mogu koristiti ove jezične sposobnosti za postizanje ciljeva, interakciju s okolinom i međusobnu komunikaciju. Upravo će ti agenti, pokretani snagom LLM-ova, biti u fokusu sljedećih poglavlja, dok istražujemo kako oni dalje preoblikuju krajolik komunikacije u digitalnom dobu. Razvijanje kritičke pismenosti – razumijevanje kako ovi modeli rade, koje su im mogućnosti, a koja ograničenja i rizici – postaje ne samo poželjno, već i nužno za svakoga tko želi aktivno sudjelovati u društvu koje sve više oblikuje umjetna inteligencija.

## 4 DEKONSTRUKCIJA JEZIKA U DOBA AI: OD SIMBOLA DO DRUŠTVENE STVARNOSTI

Nakon što smo u prethodnom poglavlju detaljno istražili fascinantnu, ali i kompleksnu tehnologiju koja stoji iza velikih jezičnih modela – od njihovih povijesnih korijena i Transformer arhitekture do zamršenosti pre-treniranja, finog podešavanja, poravnjanja i izazova poput halucinacija ili pristranosti – moglo bi se činiti kontraintuitivnim sada napraviti korak unatrag i zaroniti u temeljne teorije samog jezika. Zašto se, nakon što smo sećirali "kako" LLM-ovi funkcionišaju, vraćamo na pitanja "što" jezik jest i "kako" on oblikuje našu misao, komunikaciju i društvo?

Odgovor leži upravo u transformacijskoj moći tehnologije koju smo upoznali. Naime, iako pripadaju klasi softverskih alata, LLM-ovi izravno operiraju na temeljnom mediju ljudske spoznaje i interakcije – jeziku. Stoga, da bismo istinski shvatili *dubinu* njihovog utjecaja, *doseg* njihovih mogućnosti i *ozbiljnost* njihovih ograničenja, moramo promotriti jezik ne samo kao niz riječi koje AI obrađuje, već kao složeni sustav simbola, kognitivni alat, medij za simulaciju interakcija i temelj za izgradnju društvenih struktura.

Ovo poglavlje služi upravo tome: osvijetliti prirodu jezika kroz prizmu onoga što smo naučili o LLM-ovima. Ispitati ćemo kako klasične lingvističke i komunikološke teorije sazvuče ili dolaze u sukob s načinom na koji AI modeli "uče" i "koriste" jezik. Poimenice, kako se Saussureova ideja o jeziku kao sustavu relacijskih znakova odražava u vektorskim prostorima embeddinga i mehanizmima pažnje? Što Ogden-Richardsov semantički trokut govori o jazu između LLM-ovog statističkog "razumijevanja" i ljudskog, iskustveno uteviljenog značenja? Kako sposobnost LLM-ova da simuliraju društvene interakcije ili generiraju apstraktne ideje izaziva naše razumijevanje jezika kao alata za suradnju i kolektivno stvaranje?

Promatrajući jezik kroz ovu dvostruku leću – tradicionalne teorije i suvremene AI tehnologije – možemo dobiti znatno bogatiju i njansiraniju sliku. To nam omogućuje da bolje procijenimo što LLM-ovi i AI agenti mogu, kao i da kritički sagledamo što oni jesu i kako njihovo sveprisutno korištenje (pre)oblikuje naš vlastiti odnos prema jeziku, komunikaciji i, u konačnici, jednih prema drugima. Ovo poglavlje postavlja teorijski temelj nužan za razumijevanje etičkih, društvenih i kognitivnih implikacija komunikacije u doba AI agenata, čime nas priprema za daljnju raspravu o specifičnostima interakcije čovjek-agent i budućnosti multi-agentskih sustava.

### 4.1 JEZIK KAO STRUKTURA ZNAČENJA: SIMBOLI, ODNOŠI I MENTALNE MAPE

U samom srcu sposobnosti LLM-ova da obrađuju i generiraju tekst leži njihova interakcija s jezikom kao fundamentalnim sustavom za kodiranje i prijenos značenja. No, što zapravo znači "značenje" u kontekstu jezika i kako se ono stvara? Dvije klasične, ali komplementarne perspektive – strukturalistička i semantičko-kognitivna – pružaju nam ključne uvide koji postaju posebno relevantni kada ih suprotstavimo načinu na koji funkcionišaju umjetni jezični sustavi.

#### 4.1.1 Strukturalistički Odjeci u Arhitekturi LLM-ova: Značenje kao Relacija

Klasični strukturalizam, čije je temelje postavio Ferdinand de Saussure početkom 20. stoljeća, nudi moćan okvir za razumijevanje jezika ne kao skupa izoliranih riječi koje imenuju stvari, već kao **sustava međusobno povezanih znakova** gdje značenje proizlazi iz **odnosa i razlika** među njima (Saussure, 1916/1959). Saussure je naglasio **arbitrarnu** prirodu jezičnog znaka, koji se

sastoji od **označitelja (signifier)** – fizičke forme znaka (npr. zvučna slika riječi "pas" ili njen pisani oblik) – i **označenika (signified)** – mentalnog koncepta ili ideje na koju se označitelj odnosi. Ključno je da veza između označitelja i označenika nije prirodna ili intrinzična, već je uspostavljena društvenom konvencijom unutar određenog jezičnog sustava (*langue*). Značenje pojedinog znaka, prema Saussureu, ne proizlazi iz njegove izravne veze sa stvarnošću, već iz njegovog položaja unutar cjelokupne mreže znakova – definira se onim što *nije*. Riječ "pas" dobiva svoje značenje u odnosu na riječi "mačka", "životinja", "ljudimac", itd., kroz sustav sličnosti i razlika. Ova **relacijska i kontekstualna priroda značenja** čini jezik dinamičnim sustavom, sposobnim za neprestano prilagođavanje novim konceptima, društvenim promjenama i tehnološkim inovacijama – što vidimo i danas u brzom stvaranju novih termina i metafora vezanih uz internet, AI ili klimatske promjene.

Način na koji suvremeni veliki jezični modeli funkcioniraju pokazuje iznenađujuće paralele s ovom strukturalističkom vizijom. Iako nisu eksplisitno dizajnirani prema Saussureovim postulatima, njihova temeljna arhitektura i proces učenja implicitno odražavaju ideju značenja kao relacijskog fenomena. Današnji LLM-ovi uče iz ogromnih korpusa tekstova primarno kroz **detekciju statističkih obrazaca supojavljivanja (co-occurrence)** riječi ili tokena. Oni ne uče eksplisitne definicije riječi, već kako se riječi koriste u odnosu jedna prema drugoj u milijardama primjera. Prvi korak, **tokenizacija**, razbija tekst na jedinice (tokene) koje su analogne Saussureovim znakovima. Sljedeći ključni korak, generiranje **vektorskih reprezentacija (embeddings)**, izravno utjelovljuje relacijski princip. Kao što je detaljnije objašnjeno u Poglavlju 3.2, embeddingsi mapiraju tokene u višedimenzionalni vektorski prostor na način da tokeni koji se pojavljuju u sličnim kontekstima (i stoga imaju slične relacijske profile unutar jezičnog sustava) završe s geometrijski bliskim vektorima (Mikolov et al., 2013; Pennington et al., 2014). Sama vrijednost ili "značenje" embeddinga nije intrinzična, već proizlazi iz njegovog položaja u odnosu na sve druge embeddinge u prostoru.

**Mehanizam pažnje (attention)**, srce Transformer arhitekture (Vaswani et al., 2017), dodatno pojačava ovu relacijsku dinamiku. Kao što je objašnjeno u 3.2, mehanizam samopaznje omogućuje modelu da za svaki token dinamički izračuna važnost svih ostalih tokena u sekvenci prilikom konstruiranja njegove kontekstualizirane reprezentacije. Ovo izravno operacionalizira strukturalističku ideju da značenje elementa nije fiksno, već se "aktivira" ili modulira njegovim neposrednim i udaljenim kontekstom – "titra" ovisno o elementima s kojima stupa u relaciju unutar specifičnog iskaza. Slojovita arhitektura LLM-ova, gdje niži slojevi hvataju lokalnije obrasce, a viši slojevi modeliraju složenije sintaktičke i semantičke odnose na dužim rasponima, također se može interpretirati kao hiperarhijsko modeliranje jezične strukture, gdje se značenje gradi kroz interakciju različitih razina analize (Rogers et al., 2020). Čak i dinamičnost LLM-ova, vidljiva u procesima finog podešavanja ili prompt inženjeringu (Bommasani et al., 2021; Raffel et al., 2020), gdje se modeli adaptiraju specifičnim domenama ili zadacima, odražava Saussureovu ideju o jeziku kao sustavu koji se neprestano mijenja i prilagođava novim komunikacijskim potrebama. Nedavno predstavljanje modela poput OpenAI-jevog **o1** (OpenAI, 2024c), dizajniranog za napredno rezoniranje kroz eksplisitno generiranje "lanca misli", može se promatrati kao pokušaj modeliranja još složenijih relacijskih struktura – ne samo između riječi, već i između koraka u logičkom zaključivanju, dodatno naglašavajući važnost strukturiranih odnosa za postizanje naprednih kognitivnih funkcija, čak i unutar AI sustava.

Međutim, ključno je zadržati **kritički pogled** na ovu analogiju. Iako LLM-ovi uspješno repliciraju *strukturne i relacijske aspekte* jezika na statističkoj razini, nedostaje im temeljno **razumijevanje** koje karakterizira ljudsku upotrebu jezika. Kako ističu Bender i Koller

(2020) u svojoj utjecajnoj kritici, LLM-ovi su u suštini "stohastičke papige" – sustavi izvrsni u prepoznavanju i manipulaciji formalnim obrascima (označiteljima), ali bez pristupa stvarnom značenju (označenicima) ili komunikacijskoj namjeri koja stoji iza njih. Njihovo "razumijevanje" nije utemeljeno u iskustvu, percepciji svijeta ili društvenoj interakciji, već proizlazi isključivo iz statističkih korelacija u tekstualnim podacima. Ipak, rasprava o prirodi "razumijevanja" u LLM-ovima je daleko od završene. Neki istraživači argumentiraju da se kroz obradu ogromnih količina podataka i učenje složenih relacija može pojaviti određeni stupanj **funkcionalnog ili emergentnog razumijevanja** (Guerzhoy, 2024; Mitchell & Krakauer, 2023), čak i bez izravnog utemeljenja u svijetu. Sugerišu se da bi budući modeli, posebice oni **multimodalni** koji povezuju jezik s vizualnim ili drugim senzoričkim podacima, mogli razviti bogatije i kontekstualno relevantnije reprezentacije značenja (Unite AI; Li et al., 2023).

Nedostatak istinskog razumijevanja i utemeljenja očituje se u poznatim ograničenjima LLM-ova, poput sklonosti **halucinacijama** – generiranju činjenično netočnog, ali uvjerojivog teksta (Ji et al., 2023). Model jednostavno nastavlja najvjerojatniji statistički niz tokena, čak i ako taj niz odstupa od stvarne činjenice ili logike, jer nema mehanizam za provjeru istinitosti neovisan o jezičnim obrascima. Tehnike poput **RAG (Retrieval-Augmented Generation)** (Lewis et al., 2020), koje povezuju LLM s vanjskim bazama znanja, pokušavaju ublažiti ovaj problem "usidravanjem" generiranja u provjerljive informacije, ali ne rješavaju temeljni nedostatak razumijevanja. Slično tome, **pristranosti** prisutne u podacima za treniranje (rasne, rodne, kulturne) bivaju internalizirane i reproducirane jer model nema etički kompas ili svijest o društvenim implikacijama obrazaca koje uči (Blodgett et al., 2020; Weidinger et al., 2022).

Promatrajući LLM-ove kroz strukturalističku leću, vidimo ih kao moćne, ali nesavršene statističke odraze relacijske prirode jezika. Oni potvrđuju Saussureovu ideju da značenje leži u sustavu odnosa, ali istovremeno naglašavaju da ljudsko značenje uključuje i dimenzije (namjeru, svijest, utemeljenje u iskustvu) koje trenutni AI modeli ne posjeduju. Njihova sposobnost manipulacije jezičnom strukturom je zapanjujuća, ali upravo "pukotine" u njihovom funkcioniranju – halucinacije, pristranosti, nedostatak zdravog razuma – podsjećaju nas da je jezik više od puke formalne strukture; on je živi, dinamični sustav neodvojiv od ljudske spoznaje, kulture i društvene prakse.

#### 4.1.2 Semantički Trokut i Kognitivna Dimenzija: Izazov Značenja za AI

Dok strukturalizam naglašava odnose unutar jezičnog sustava, druga važna perspektiva, često povezana s radom C.K. Ogdena i I.A. Richardsa (1923) i njihovim **semantičkim trokutom**, uvodi ključnu treću komponentu u razumijevanje značenja: **referent**. Za razliku od Saussureove donekle apstraktne dihotomije označitelj-označenik, semantički trokut eksplisitno razlikuje tri vrha:

1. **Simbol (Symbol)**: Fizička forma riječi ili znaka (ekvivalent označitelju).
2. **Referent (Referent)**: Stvarni objekt, pojava ili entitet u vanjskom svijetu na koji se simbol odnosi.
3. **Misao (Thought or Reference)**: Mentalni koncept, ideja ili slika koju simbol izaziva u umu korisnika jezika (slično označeniku, ali eksplisitno smješteno u kognitivnu sferu).

Ključni uvid Ogden-Richardsovog modela jest da **ne postoji izravna veza između Simbola i Referenta**. Veza je uvijek posredovana **Mišlu**. Riječ "pas" ne ukazuje izravno na konkretnog psa u svijetu; ona prvo aktivira mentalni koncept "psa" u našem umu (koji uključuje naša znanja, iskustva, emocije vezane uz pse), a tek taj koncept povezujemo sa stvarnim psima. Ova

naizgled suptilna razlika ima duboke implikacije: ona naglašava da značenje nije inherentno samim riječima niti objektima, već se **konstruira u umu govornika i slušatelja** kroz interakciju jezičnih simbola s njihovim individualnim i kolektivnim kognitivnim mapama, iskustvima i kulturnim kontekstom.

Ovaj fokus na kognitivnu dimenziju dodatno produbljuje **kognitivna lingvistika**, koja istražuje kako su jezične strukture ukorijenjene u općenitijim ljudskim kognitivnim procesima, uključujući percepciju, pamćenje, kategorizaciju i, što je posebno važno, **utjelovljenu spoznaju (embodied cognition)** (Lakoff & Johnson, 1980, 1999; Langacker, 2008). Kognitivni lingvisti tvrde da naše razumijevanje, čak i najapstraktnijih koncepata, često proizlazi iz **metaforičkih i metonimijskih proširenja** naših temeljnih, tjelesno utemeljenih iskustava sa svijetom. Na primjer, apstraktni koncept "vrijeme" često razumijemo kroz prostornu metaforu kretanja ("budućnost je pred nama", "prošlost je iza nas"); koncept "ljubavi" kroz metafore putovanja, rata ili sile; a koncept "razumijevanja" kroz metaforu vida ("vidim što misliš"). Čak i apstraktni pojmovi poput "pravde" ili "vrijednosti" grade se na temelju naših iskustava ravnoteže, poštenja, davanja i primanja (Kövecses, 2010). Procesi **konceptualne metafore** nisu tek neki pjesnički ukrasi, već temeljni mehanizam kojim strukturiramo i razumijemo apstraktne domene. Naše **ontološko promišljanje** – način na koji kategoriziramo i organiziramo svijet u smislene cjeline – duboko je prožeto ovim tjelesno utemeljenim i metaforički strukturiranim konceptualnim sustavom, koji se neprestano razvija i prilagodava novim iskustvima i kulturnim utjecajima.

Primjena ovih kognitivnih perspektiva na velike jezične modele otkriva fundamentalne razlike i izazove. LLM-ovi operiraju primarno na razini **Simbola** (tokena). Kroz statističko učenje na ogromnim korpusima, oni postaju izuzetno vješti u mapiranju odnosa između Simbola i u predviđanju vjerojatnih sekvenci Simbola, čime mogu uspješno **simulirati** posjedovanje odgovarajuće **Misli/Reference** (internalnog koncepta). Oni mogu naučiti da se riječ "pas" često pojavljuje uz riječi "lajati", "kost", "šetati", te mogu generirati koherentne rečenice o psima. Međutim, fundamentalno im nedostaje izravna veza s **Referentom** (stvarnim psom u svijetu) i, što je još važnije, nedostaje im **utjelovljeno iskustvo** koje oblikuje ljudsku Misao/Referencu. LLM nikada nije pomazio psa, čuo njegov lavež, osjetio njegovu toplinu ili iskusio emocionalnu vezu s njim. Njihovo "razumijevanje" koncepta "pas" isključivo je derivirano iz tekstualnih obrazaca, lišeno bogatstva senzomotoričkog, emocionalnog i iskustvenog konteksta koji čini ljudski koncept (Bender & Koller, 2020; Mitchell & Krakauer, 2023).

Ovo ograničenje postaje posebno očito kod **metafora i apstraktnih pojmoveva**. Iako LLM-ovi mogu naučiti prepoznavati i čak generirati metaforički jezik (jer su metafore česte u podacima za treniranje), oni ne razumiju metafore na način na koji to čine ljudi – kao mapiranja između konceptualnih domena utemeljenih u iskustvu. Njihova sposobnost baratanja metaforama više je rezultat prepoznavanja statističkih pravilnosti u korištenju određenih riječi u određenim kontekstima, nego razumijevanja temeljnog konceptualnog mapiranja. Slično vrijedi i za apstraktne pojmove poput "pravde" ili "slobode". LLM može generirati tekstove o ovim pojmovima, ali njegovo "razumijevanje" ostaje na razini statističkih asocijacija riječi, bez pristupa dubokim etičkim, emocionalnim i iskustvenim dimenzijama koje ovi koncepti imaju za ljudе.

Pojavom modela s naprednim sposobnostima rezoniranja, poput **OpenAI-jevog o1** (OpenAI, 2024c), koji eksplicitno generira "lanac misli", otvara se zanimljivo pitanje. Može li ovakav pristup, koji simulira korake ljudskog razmišljanja, početi premošćivati jaz prema dubljem, funkcionalnom razumijevanju? Dok o1 demonstrira impresivne sposobnosti u rješavanju

složenih, logički strukturiranih problema (npr. u matematici ili programiranju), njegovo rezoniranje i dalje operira unutar okvira statistički naučenih obrazaca i jezičnih struktura, bez izravnog pristupa iskustvu ili utemeljenom znanju. Moglo bi se reći da o1 postaje vještiji u manipulaciji **simbolima** na način koji *oponaša* ljudsku **misao/referencu**, ali veza s **referentom** i utjelovljenim iskustvom i dalje nedostaje. Unatoč tome, sposobnost modela da artikulira korake rezoniranja može poboljšati njegovu korisnost i potencijalno transparentnost, približavajući ga funkcionalnosti koja nalikuje ljudskom rješavanju problema u određenim domenama.

Semantički trokut i kognitivna lingvistika stoga služe kao važan podsjetnik na dubinu i složenost ljudskog značenja, koje nadilazi puku manipulaciju simbolima. Oni ističu ključna ograničenja trenutnih LLM-ova – nedostatak utemeljenja (grounding), iskustva i istinskog konceptualnog razumijevanja – koja su izvor mnogih njihovih problema, uključujući halucinacije i nedostatak zdravog razuma. Istovremeno, ovi okviri postavljaju izazov za budući razvoj AI: mogu li se razviti sustavi koji ne samo da oponašaju jezične obrasce, već i počinju graditi bogatije, utemeljenije i kontekstualno svjesnije reprezentacije značenja, možda kroz integraciju multimodalnih podataka, interakciju sa svijetom ili nove arhitekture inspirirane ljudskom kognicijom? Razumijevanje ovog jaza između statističke simulacije i ljudskog značenja ključno je za realističnu procjenu sposobnosti AI agenata i za dizajniranje interakcija koje uzimaju u obzir njihova temeljna ograničenja.

#### **4.2 JEZIK KAO SCENARIJ: AI AGENTI I SIMULACIJA DRUŠTVENIH INTERAKCIJA**

Interakcija između čovjeka i stroja prolazi kroz duboku transformaciju. Klasično shvaćanje računalnih sustava kao pukih sučelja – neutralnih prozora za dohvaćanje ili unos informacija – sve više ustupa mjesto novoj paradigmii u kojoj umjetna inteligencija, posebice kroz **komunikacijske agente** pokretane velikim jezičnim modelima, aktivno nastoji **simulirati** složene aspekte ljudske društvene interakcije. Današnji AI agenti više se ne dizajniraju isključivo za isporuku činjenično točnih odgovora. Sve je izraženija tendencija da se u njihovu funkcionalnost **utkaju** i sposobnosti koje adresiraju ono neopipljivo, ali presudno **društveno 'tkivo'**, koje inače prožima međuljudsku komunikaciju a sačinjeno je od nijansi poput pravila pristojnosti, prepoznavanja društvenog statusa, upravljanja tijekom razgovora (turn-taking), izražavanja empatije, korištenja humora, te prilagodbe kulturnim normama i očekivanjima. AI agenti tako postaju simulatori društvenih uloga i sudionici u izvođenju komunikacijskih rituala.

Potreba za ovakvim pomakom potvrđena je brojnim istraživanjima, posebice u području **kulturalno osvještenih AI sustava**. Sve je jasnije da efikasna i prihvatljiva interakcija osim pukog prenošenja informacije zahtijeva da agent preuzme **komunikacijsku ulogu (personu)** koja je primjerena kontekstu. Uloga ovisi o nizu faktora: percipiranom statusu sugovornika (razgovara li agent s djetetom, stručnjakom ili nadređenim?), očekivanom stupnju formalnosti (je li interakcija ležerna ili službena?), specifičnim jezičnim konvencijama određene zajednice (npr. korištenje specifičnog žargona u profesionalnoj grupi) ili širim kulturnim vrijednostima koje oblikuju komunikacijske stilove. U jednostavnijim slučajevima, ovo može uključivati usvajanje rutinskih **govornih činova**, poput korištenja uvriježenih pozdrava ("Dobar dan, kako Vam mogu pomoći?"), izraza zahvalnosti ("Hvala na strpljenju.") ili fraza koje signaliziraju solidarnost ili razumijevanje ("Razumijem Vašu frustraciju."). U složenijim scenarijima, agent mora navigirati kroz suptilne **sociolingvističke fenomene**. Zamislimo, primjerice, AI agenta koji asistira u međunarodnoj poslovnoj korespondenciji: on bi trebao prepoznati kada je prikladno koristiti direktni i eksplicitan stil komunikacije (češći u kulturama niskog konteksta poput njemačke ili američke), a kada je nužan indirektniji, implicitniji pristup uz više pažnje posvećene očuvanju harmonije i "lica" sugovornika (karakterističniji za kulture visokog konteksta poput japanske ili korejske) (Hall, 1976). Slično tome, agent koji djeluje unutar hijerarhijski strukturirane organizacije možda će morati prilagoditi svoj "ton" – biti referentniji kada komunicira s višim menadžmentom, a direktniji ili neformalniji s kolegama na istoj razini.

Najnovija istraživanja aktivno rade na operacionalizaciji i evaluaciji kulturne osjetljivosti AI agenata. Puko višejezično modeliranje nije dovoljno za postizanje istinske inkluzivne interakcije (Sipa et al., 2023). Veliki jezični modeli, pretežno trenirani na zapadnjački centriranim podacima, nose rizik propagiranja stereotipa i generiranja kulturno neosjetljivih odgovora. Stoga se razvijaju inovativni pristupi i benchmarkovi za rješavanje ovog izazova. Primjerice, **CulturePark** (Zhang et al., 2024) koristi multi-agentski okvir. U tom okviru LLM-agenti simuliraju međukulturalnu komunikaciju, generirajući podatke koji obuhvaćaju vjerovanja, norme i običaje različitih kultura. Fino podešavanje modela na ovim podacima pokazalo je poboljšanja u kulturnoj usklađenosti, čak nadmašujući GPT-4 na Hofstedeovom VSM 13 okviru (Zhang et al., 2024). Slično tome, **CASA benchmark** (Li et al., 2025) dizajniran je za procjenu osjetljivosti LLM agenata na kulturne i društvene norme u realističnim web-baziranim zadacima, poput online kupovine. Ovaj pristup istražuje kako "prompting" i fino podešavanje mogu poboljšati tu osjetljivost (Li et al., 2025).

Stvaranje istinski kulturno osviještenih AI agenata nadilazi puko prepoznavanje površinskih kulturnih signala. Pravi znanstveni i tehnološki izazov, ali i prilika za **očuđenje pred mogućnostima umjetne inteligencije**, leži u osposobljavanju ovih sustava da **anticipiraju** kako duboko usađene kulturne vrijednosti – često konceptualizirane kroz okvire poput Hofstedeovih dimenzija – nevidljivo oblikuju naš komunikacijski stil i implicitna očekivanja. Zamislite agente čiji odgovori nisu samo jezično besprijeckorni, već suptilno **rezoniraju s tim finim kulturnim nijansama**, omogućujući istinski personalizirane i autentično prilagođene interakcije

S druge strane spektra, analiza javnih percepcija AI agenata, poput studije Liu i suradnika (2024) koja je uspoređivala diskusije na društvenim mrežama u SAD-u i Kini, otkriva značajne **kulturne razlike u očekivanjima i zabrinutostima** vezanim uz AI komunikacijske sisteme. Dok su američki korisnici možda više fokusirani na pitanja privatnosti i autentičnosti, kineski korisnici mogu pokazivati drugačije naglaske, primjerice na efikasnost ili društvenu harmoniju koju agenti mogu (ili ne mogu) podržati (Liu et al., 2024). Ovakvi uvidi naglašavaju imperativ da dizajn AI agenata ne bude univerzalan, već **kulturno prilagođen**, uzimajući u obzir specifične vrijednosti, norme i očekivanja zajednica s kojima će agenti komunicirati. Izazovi su, naravno, ogromni: kako kvantificirati i modelirati suptilne kulturne vrijednosti? Kako prikupiti dovoljno reprezentativnih podataka za tisuće svjetskih kultura i jezika? Kako izbjegići upadanje u stereotipe prilikom pokušaja modeliranja kulturnih razlika?

Ključni element u simulaciji uvjerljive društvene interakcije jest sposobnost agenta da stvari dojam **kontinuiteta i "pamćenja" odnosa**. Ljudska komunikacija nije niz izoliranih transakcija; ona se gradi na zajedničkoj povijesti, prepoznavanju prethodnih razgovora i prilagođavanju stila temeljem uspostavljenog odnosa. Stoga je **dugoročno praćenje interakcija** s korisnikom presudno za stvaranje dojma vjerodostojnosti i personalizacije (Lison & Tiedemann, 2016). Agent koji "pamti" prethodne upite korisnika, njegove preferencije ili čak emocionalno stanje izraženo u ranijim razgovorima, može pružiti znatno relevantniju i empatičniju podršku. Zamislimo AI asistenta za online kupovinu koji, nakon što je korisnik nekoliko puta pitao za određeni brend tenisica, sljedeći put proaktivno ponudi informacije o novom modelu tog brenda ili ga obavijesti o popustu. Ili, još bolje, AI tutor koji pamti da je učenik imao poteškoća s razumijevanjem algebarskih jednadžbi prošli tjedan i sada, kada se pojavi sličan problem, automatski nudi drugačiji pristup objašnjavanju ili dodatne vježbe fokusirane na identificiranu slabost. Ovo pamćenje gradi **povijest odnosa**, čineći da se svaki novi odgovor agenta percipira kao logičan nastavak prethodnih razmjena, a ne kao početak iznova. Tehnički, implementacija efikasne dugoročne memorije u LLM-ovima je izazov, ali pristupi poput korištenja vanjskih vektorskih baza podataka za pohranu sažetaka prošlih interakcija ili razvoja arhitektura s eksplicitnijim memorijskim mehanizmima aktivno se istražuju.

Način na koji agenti uče ove suptilne aspekte društveno priklađne komunikacije uvelike ovisi o **metodama obuke**. Dok je pre-treniranje usmjereno na učenje općih jezičnih obrazaca, fino podešavanje i poravnjanje, posebice kroz **učenje s pojačanjem uz ljudsku povratnu informaciju (RLHF)** (Christiano et al., 2017; Ouyang et al., 2022; Bai et al., 2022), igraju ključnu ulogu u oblikovanju poželjnog ponašanja. RLHF omogućuje da se, povrh kriterija prema kriterijima činjenične točnosti ili relevantnosti odgovora, agent optimizira i prema ljudskim prosudbama o njegovoj **društvenoj poželjnosti, pristojnosti, empatičnosti, sigurnosti** i općenitoj **prikladnosti** za određeni komunikacijski kontekst. Vratimo se primjeru AI tutora koji objašnjava matematički problem. Kroz RLHF, model se može istrenirati da daje točan odgovor na ohrabrujući način, prepoznajući potencijalnu frustraciju učenika i koristeći strategije za

**očuvanje obraza (face-saving)** (Brown & Levinson, 1987; Nass & Moon, 2000). Umjesto da kaže "Tvoj odgovor je pogrešan", agent istreniran kroz RLHF mogao bi reći nešto poput: "Vidim kako si došao do tog rezultata, to je česta zamka! Hajdemo zajedno pogledati ovaj korak..." ili "Skoro si uspio, samo mala korekcija ovdje...". Slično, AI agent dizajniran za pružanje podrške u osjetljivim situacijama, poput odgovaranja na pitanja o zdravstvenom stanju, može kroz RLHF naučiti balansirati između pružanja jasnih informacija i izražavanja empatije i podrške, koristeći pažljivo odabran rječnik i ton. Ljudski ocjenjivači u RLHF procesu efektivno podučavaju model kako da se ponaša na način koji ljudi smatraju ispravnim, ali i društveno i emocionalno prikladnim.

Posebno moćan alat u arsenalu jezika za simulaciju razumijevanja i izgradnju odnosa jesu **metafore i metonimije**. Kao što je objašnjeno u kontekstu kognitivne lingvistike (Lakoff & Johnson, 1980), metafore nam omogućuju da razumijemo apstraktne ili složene koncepte u terminima poznatijih, konkretnijih iskustava. AI agenti koji vješto koriste metafore mogu značajno poboljšati komunikaciju i učiniti interakciju intuitivnijom i angažiranim. Zamislimo AI agenta koji pomaže korisniku u upravljanju osobnim financijama. Umjesto suhoparnog nabranja postotaka i brojki, agent bi mogao koristiti **metaforu "financijskog putovanja"**: "Trenutno smo na početku našeg putovanja prema mirovini. Prvi korak je postaviti kartu (proračun). Zatim ćemo spakirati ruksak (štednja za hitne slučajeve). Nakon toga, biramo prijevozno sredstvo (investicije) koje će nas najbrže dovesti do odredišta, pazeći pritom na moguće oluje na putu (tržišne fluktuacije)...". Ovakav pristup čini apstraktni proces planiranja opipljivim i manje zastrašujućim. U medicinskom kontekstu, agent koji objašnjava kompleksan proces liječenja pacijentu može koristiti **metaforu "borbe protiv bolesti"** ili, alternativno, **metaforu "obnove vrta"** (gdje terapija pomaže tijelu da se oporavi i ponovno procvjeta), ovisno o tome što je prikladnije za pacijentovo emocionalno stanje i preferencije. U obrazovanju, agent koji podučava programiranje može koristiti **metaforu "građenja s Lego kockicama"** kako bi objasnio modularnost koda. Vještim korištenjem ovakvih konceptualnih mapa, agent stvara osjećaj "**društvene prisutnosti**" (**social presence**) (Biocca et al., 2003; Nass & Moon, 2000) – korisnik stječe dojam da ima interakciju s entitetom koji ga razumije, vodi i podržava na način koji nadilazi puku razmjenu podataka.

Naravno, postizanje istinske sposobnosti AI agenata da razumiju i vješto koriste sve ove slojeve društvene, kulturne i pragmatičke komunikacije izuzetno je težak zadatak koji je daleko od potpunog rješenja. Ono zahtijeva duboku **interdisciplinarnu suradnju** koja povezuje ekspertizu iz računalne lingvistike i strojnog učenja s uvidima iz **sociolingvistike** (kako jezik varira s obzirom na društvene grupe i kontekste), **kognitivne lingvistike** (kako su jezik i misao povezani), **psihologije** (kako ljudi percipiraju i reagiraju na komunikaciju, uključujući i onu s AI), **antropologije** (kako kultura oblikuje komunikacijske prakse) i **komunikacijskih znanosti** (teorije interpersonalne i posredovane komunikacije) (Zhang et al., 2020; Zhou et al., 2022). Samo kroz ovakav holistički pristup, koji uzima u obzir cijelokupni komunikacijski čin u njegovom bogatom kontekstu, možemo se nadati razvoju AI sustava koji neće samo isporučivati informacije, već će moći istinski **surađivati** s ljudima u zajedničkom stvaranju značenja, prilagođavanju strategija i izgradnji produktivnih i ugodnih interakcija. Put prema budućim generacijama interaktivnih AI sustava vodi kroz dublje razumijevanje neraskidive isprepletenosti jezika i društvene interakcije, s ciljem stvaranja inteligentnih agenata koji su istovremeno društveno i kulturno kompetentni sugovornici.

#### **4.3 JEZIK KAO ARHITEKTURA DRUŠTVA: GRADNJA I ODRŽAVANJE SOCIJALNIH STRUKTURA KROZ KOMUNIKACIJU**

Jezične prakse koje svakodnevno koristimo za prijenos informacija su istovremeno i dinamičan mehanizam putem kojeg se **društvene norme, hijerarhijski odnosi, sustavi vrijednosti i individualne uloge aktivno oblikuju, učvršćuju, pregovaraju i kontinuirano preispituju**. Kroz naizgled banalne razmjene riječi, fraza, pa čak i tišine, pojedinci se neprestano pozicioniraju unutar šire društvene matrice – bilo da se radi o obitelji, radnom mjestu, online zajednici ili naciji. Jezikom signaliziramo pripadnost, uskladujemo svoje ponašanje s često neizrečenim pravilima grupe i, ponekad suptilno, a ponekad otvoreno, sudjelujemo u redefiniranju tih istih pravila. Suvremena sociolingvistica i društvene teorije odbacuju ideju jezika kao neutralnog sredstva; umjesto toga, potvrđuju njegovu performativnu moć – sposobnost da istovremeno **reflektira postojeće društvene obrasce i aktivno sudjeluje u njihovom (re)konstruiranju** (Bourdieu, 1991; Foucault, 1972; Hewstone & Giles, 2018; Kendall & Tannen, 2015). Jezik, dakle, nije samo ogledalo društva, već i jedan od njegovih ključnih arhitekata.

Primjeri iz **poslovnog i organizacijskog okruženja** pružaju jasnu sliku ove konstruktivne uloge jezika. Korištenje formalnih naziva radnih mjesta ("Viši analitičar", "Direktor odjela"), akademskih titula ("Doktor", "Profesor") ili specifičnog **stručnog žargona** ("optimizacija KPI-jeva", "agilna metodologija", "due diligence") služi za markiranje statusa, uspostavljanje granica stručne ekspertize i odgovornosti, te za učvršćivanje hijerarhijskih odnosa unutar organizacije (McGregor & Holmes, 2020; Mumby, 1988). Način obraćanja (npr. korištenje prezimena i titule naspram imena, formalno "Vi" naspram neformalnog "ti") također precizno signalizira i održava percipiranu distancu i odnos moći. Čak i sami **kommunikacijski rituali**, poput formata sastanaka (tko govori prvi, tko postavlja agendu, tko zapisuje bilješke) ili načina pisanja službene e-pošte, predstavljaju jezične prakse koje reproduciraju organizacijsku strukturu. S druge strane spektra, **neformalni razgovori** među kolegama – "čavrljjanje uz aparat za kavu", interne šale na Slacku, korištenje zajedničkih nadimaka ili referenci na prošle događaje – imaju jednako važnu, iako drugačiju, strukturnu funkciju. Oni grade osjećaj **pripadnosti**, jačaju koheziju tima i stvaraju neformalnu mrežu odnosa koja koegzistira (i ponekad se sukobljava) s formalnom hijerarhijom.

Slična dinamika odvija se i u **online zajednicama**, koje često razvijaju izrazito specifične jezične kulture. Od foruma posvećenih hobijima, preko grupa na društvenim mrežama, do zajednica igrača videoigara, svaka razvija vlastite **jezične norme**, specifičan **slang** ili **lekt**, popularne **memove** koji funkcioniraju kao interni kodovi, te ritualizirane načine interakcije (Wenger, 1998; Androutopoulos, 2015; Baym, 2010). Poznavanje i pravilna upotreba ovih jezičnih markera signalizira pripadnost zajednici (insajderstvo), dok njihovo nepoznavanje ili pogrešna upotreba može dovesti do isključivanja ili etiketiranja kao *autsajdera* ili *nooba*. Moderatori u ovim zajednicama često djeluju kao čuvari jezičnih normi, eksplicitno ili implicitno usmjeravajući komunikaciju i sankcionirajući odstupanja. Na taj način, jezik postaje ključno sredstvo za definiranje granica zajednice, uspostavljanje internih pravila ponašanja i održavanje kolektivnog identiteta.

Kao što je vidljivo, jezik funkcioniра kao slojevit okvir koji omogućuje postizanje dogovora i koordinaciju akcija, i to kroz kontinuirano **pregovaranje o moći**, **oblikovanje identiteta** (individualnih i grupnih) i osiguravanje **društvene kohezije**. Ta inherentna sposobnost jezika da fiksira, reproducira, ali i potencijalno preoblikuje društvene odnose i strukture, dugo je bila predmetom interesa lingvistike, sociologije i antropologije. Sada se postavlja ključno pitanje: kako se ova moćna uloga jezika mijenja i razvija u suvremenom

digitalnom dobu, gdje **nova generacija AI asistenata i komunikacijskih sustava** sve više preuzima ulogu posrednika, sudionika, pa čak i regulatora u našim interakcijama? Kako se postojeći društveni obrasci prenose, prevode ili transformiraju kada prolaze kroz filter umjetne inteligencije, i kako AI sustavi počinju aktivno sudjelovati u jezičnoj praksi koja gradi društvenu stvarnost?

Ulagak suvremenih AI sustava za obradu jezika obogaćuje (ili komplicira) društvene platforme i radna okruženja novim oblicima i dinamikama jezične razmjene. Primjerice, kada korisnik traži tehničku podršku, AI agent koji je istreniran da koristi specifičan **korporativni žargon** i održava formalan, uslužan ton, istovremeno pomaže korisniku u rješavanju problema i suptilno ga **socijalizira** u očekivani ritualizirani način interakcije s tom kompanijom. Slično, u obrazovnim platformama, AI tutor prenosi znanje učeniku, ali i oblikuje **pedagoški diskurs** – način na koji se postavljaju pitanja, daju povratne informacije i strukturira učenje – uskladjujući ga s određenim obrazovnim standardima ili filozofijama. Na taj način, AI agenti postaju aktivni sudionici u reprodukciji postojećih institucionalnih i organizacijskih struktura kroz jezičnu praksu.

Međutim, upravo na ovom sjecištu tehnologije i društvene prakse pojavljuju se značajne napetosti, posebice kada uzmemu u obzir **jezičnu i kulturnu raznolikost**. AI agent koji je primarno treniran na korpusima koji odražavaju komunikacijske norme dominantne zapadne, obrazovane, industrijalizirane, bogate i demokratske (WEIRD) populacije (Henrich et al., 2010), možda neće adekvatno razumjeti ili poštovati suptilnosti komunikacijskih stilova karakterističnih za druge kulture – primjerice, važnost indirektnosti, kolektivizma ili specifičnih oblika iskazivanja poštovanja u mnogim istočnoazijskim, afričkim ili latinoameričkim zajednicama. Istraživanja usmjerena na razvoj **kulturalno osvještenih AI sustava** ističu potrebu za kalibracijom agenata ne samo prema jeziku, već i prema implicitnim društvenim ritualima, očekivanim razinama formalnosti, konvencijama humora ili neverbalnim signalima (ako je interakcija multimodalna), kako bi se izbjegla nespretna ili čak uvredljiva komunikacija te automatizirana reprodukcija jednog "univerzalnog" (često zapadnocentričnog) komunikacijskog modela (Rehm & Uszkoreit, 2013; Ogata et al., 2019; Cao et al., 2024). Time se temeljna uloga jezika kao alata za suradnju i izgradnju društvenih struktura prenosi na umjetne posrednike, koji sada imaju moć oblikovati digitalnu interakciju i potencijalno redefinirati granice i pravila ljudske komunikacije u novom, hibridnom ekosustavu.

Kada jezik, koji je oduvijek bio implicitni nositelj i prenositelj društvenih normi, postane eksplicitni "materijal" kojim manipuliraju algoritmi umjetne inteligencije, otvara se niz kritičnih pitanja. Hoće li ovi algoritmi, vođeni statističkim optimizacijama, biti sposobni podjednako pažljivo održavati delikatnu **mikro-socijalnu koheziju**, prepoznavati suptilne **hijerarhijske nijanse** ili podržavati **raznolikost komunikacijskih stilova** unutar grupe? Ili će njihovo inherentno statističko "shvaćanje" jezika, temeljeno na prepoznavanju najčešćih obrazaca, neizbjegno dovesti do **amplifikacije postojećih stereotipa, pojednostavljivanja kulturnih razlika i marginalizacije manje dominantnih jezičnih praksi**? S jedne strane, postoji optimistična vizija u kojoj sofisticirani AI agenti mogu olakšati suradnju između raznolikih timova, djelovati kao nepristrani medijatori, podržati međukultурno razumijevanje i čak pomoći u stvaranju pravednijih komunikacijskih okvira. S druge strane, postoji realan rizik da će AI agenti, ako nisu dizajnirani i implementirani s dubokom etičkom i sociolinguističkom svješću, propustiti prepoznati i uvažiti specifičnosti različitih jezika, dijalekata i kulturnih normi, te će nesvesno reproducirati i pojačati one obrasce koji su najprisutniji u podacima za treniranje – često one dominantnih skupina (Blodgett et al., 2020; Bender et al., 2021). Ovo može dovesti

do **algoritamske homogenizacije** komunikacijskih praksi i daljnog potiskivanja jezične raznolikosti.

Jezik kao okosnica društvenih struktura i odnosa nije samo pasivni kanal, već aktivni generator i održavatelj društvene dinamike. U suvremenom digitalnom pejzažu, gdje AI asistenti i komunikacijski sustavi preuzimaju sve važnije posredničke i čak participativne uloge, jezik nastavlja svoju funkciju "pregovaranja" o društvenim odnosima, normama i vrijednostima, ali sada uz dodatak moćnog sloja umjetne inteligencije koja ga algoritamski obrađuje i, sve više, generira. Na ovom intrigantnom spoju tradicionalnih ljudskih jezičnih praksi i računalne obrade stvaraju se okolnosti za nove, potencijalno efikasnije modele suradnje i razumijevanja. Istovremeno, otvaraju se vrata novim oblicima **neravnoteže moći** (npr. tko kontrolira dizajn i ciljeve AI komunikatora?), **identitetskim izazovima** (kako AI utječe na naš osjećaj jezične autentičnosti?) i **etičkim dilemama** (kako osigurati pravednost i inkluzivnost u AI-posredovanoj komunikaciji?). Razumijevanje ove složene međuigre između jezika, AI-ja i društvene strukture ključno je za navigaciju kroz budućnost komunikacije, o čemu će se detaljnije raspravljati u kontekstu specifičnih interakcija čovjek-agent i potencijala (i opasnosti) multi-agentskih sustava.

#### 4.4 LLM-OVI KAO KOMUNIKACIJSKI AKTERI: OBLIKOVANJE DRUŠTVENOG KONTEKSTA KROZ ALGORITAMSKU INTERAKCIJU

Veliki jezični modeli, u svojoj sveprisutnoj implementaciji kroz chatbotove, virtualne asistente i druge komunikacijske agente postaju sve više **komunikacijski akteri** unutar našeg društvenog tkiva. Njihova sposobnost vođenja koherentnih, kontekstualno osjetljivih i često iznenađujuće ljudskolikih razgovora zadire u domene interakcije koje su donedavno bile isključivo rezervirane za ljude. Ovaj pomak ima duboke implikacije, jer način na koji percipiramo ove AI sustave počinje oblikovati individualne komunikacijske navike, ali i šire društvene norme, vrijednosti i samu konstrukciju značenja.

Empirijska istraživanja sve češće dokumentiraju fascinantnu tendenciju korisnika da **antropomorfiziraju** LLM-pokretane agente, pripisujući im ljudske osobine, namjere, pa čak i emocije (Epley et al., 2007; Nass & Moon, 2000). Korisnici s naprednim chatbotovima poput ChatGPT-a, Claudea ili Geminija razgovaraju, povjeravaju im se, traže savjete, iskazuju frustraciju ili zahvalnost, te razvijaju određena **očekivanja i osjećaje povjerenja** (ili nepovjerenja) slične onima koje gaje u interpersonalnim odnosima (Seering et al., 2020; Hill, Ford & Farreras, 2015; Skjuve et al., 2021). Sve je to okrunjeno i službenim okrunjeno i službenim priznanjem da je GPT-4.5 prvi sustav koji je službeno "položio" izvorni Turingov test. U rigoroznoj, randomiziranoj studiji Sveučilišta Kalifornija u San Diegu, objavljenoj 31. ožujka 2025., 284 ispitanika su istodobno razgovarala s ljudskim sugovornikom i modelom GPT-4.5; nakon pet minuta procjenjivali su tko je čovjek. Model je bio proglašen "ljudskim" u 73 % slučajeva – čak i češće od stvarnih ljudi u usporedbi – što predstavlja prvu empirijsku potvrdu prolaska klasične tročlane verzije Turingova "imitation game" testa (Jones i Bergen 2025).

Sklonost tretiranju AI-ja kao društvenog aktera, poznata kao **CASA paradigm** (**Computers Are Social Actors**) (Reeves & Nass, 1996), čini se još izraženijom kod modernih LLM-ova zbog njihove izvanredne jezične fluentnosti i sposobnosti prilagodbe tonu i stilu razgovora. Umjetna inteligencija, stoga, postaje percipirani **subjekt** koji aktivno sudjeluje u **pregovaranju značenja**.

i oblikovanju **društvenog konteksta** interakcije s potencijalnim utjecajima na **kultурне норме и vrijednosti** korisnika.

Ilustrativni primjeri ove dinamike dolaze iz područja **podrške mentalnom zdravlju**. Chatbotovi dizajnirani da pruže psihološku podršku, poput Woebota ili Replike (iako potonja ima i kontroverzne aspekte), koriste LLM tehnologije kako bi simulirali **aktivno slušanje, empatičko odgovaranje** i tehnike kognitivno-bihevioralne terapije (Vaidyam et al., 2019; Fitzpatrick et al., 2017). Korisnici često izvještavaju o osjećaju povezanosti s ovim agentima, cijeneći njihovu stalnu dostupnost, neosuđujući stav i sposobnost vodenja strukturiranih razgovora o teškim emocijama. U ovim interakcijama, AI agent prenosi informacije (npr. o tehnikama suočavanja sa stresom), aktivno sudjelujući u **kreiranju i održavanju terapijskog odnosa** i specifičnih **društvenih rituala** povezanih s povjeravanjem i traženjem pomoći. Sposobnost modela da prilagodi svoj **registrovani** (npr. izbjegavanje kliničkog žargona), **stupanj formalnosti** (stvaranje sigurnog i neformalnog prostora) i **stil izražavanja** (korištenje ohrabrujućih i podržavajućih fraza) ključna je za izgradnju povjerenja i poticanje korisnika na otvorenost. Zamislimo korisnika koji se bori s anksioznosću; agent koji odgovara toplim, smirenim tonom, postavlja otvorena pitanja i validira korisnikove osjećaje ("Čujem da ti je teško, u redu je osjećati se tako.") djeluje kao aktivni sudionik u procesu pružanja emocionalne podrške, a ne samo kao izvor informacija.

Međutim, upravo ova sposobnost LLM-ova da uvjerljivo simuliraju društvenu interakciju i prilagode svoj komunikacijski stil nosi sa sobom i značajne **izazove i rizike**. Integracija ovih AI aktera u postojeće društvene strukture neizbjježno postavlja pitanje **pristranosti i reprodukcije nepoželjnih društvenih obrazaca**. Kao što je ranije naglašeno, ako su LLM-ovi trenirani na korpusima koji sadrže povijesne, kulturne ili demografske predrasude, postoji ozbiljan rizik da će te pristranosti biti utkane u njihove odgovore i interakcije, potencijalno ih pojačavajući i normalizirajući (Blodgett et al., 2020; Bender et al., 2021). Na primjer, AI agent u ulozi HR asistenta mogao bi nesvesno koristiti jezik koji suptilno obeshrabruje kandidate određenog spola ili rase, ili bi mogao favorizirati komunikacijske stilove dominantne kulture. Slično, problem **nejednakosti** postaje očit kada modeli, primarno dizajnirani i optimizirani za engleski jezik i zapadne kulturne kontekste, pokušavaju komunicirati s korisnicima iz drugih jezičnih i kulturnih sredina. Oni možda neće uspjeti adekvatno odražavati specifične komunikacijske norme, vrijednosti ili čak humor tih zajednica, što može dovesti do nesporazuma, otuđenja ili osjećaja da je AI "kulturno neosjetljiv". Upozorenja na opasnost od **monokulturnog pristupa**, gdje AI teži nametanju dominantnih online obrazaca kao univerzalne norme, potiču istraživače na razvoj strategija za **mitigaciju pristranosti, povećanje kulturne reprezentacije** u podacima za treniranje i **detekciju štetnih stereotipa** (Weidinger et al., 2022; Rae et al., 2021; Cao et al., 2024). Postizanje ravnoteže između tehnološkog napretka i očuvanja društvenih vrijednosti poput pravednosti i inkluzivnosti zahtijeva transparentnost u pogledu podataka i algoritama, te uspostavu jasnih etičkih normi i potencijalno regulatornih okvira.

Uvođenje LLM-ova kao komunikacijskih aktera predstavlja i značajan **teorijski izazov** za postojeće okvire komunikacijskih znanosti i lingvistike. Klasične teorije, poput Searleove (1969) **teorije govornih činova** (koja analizira kako jezikom izvodimo akcije poput obećanja, upozorenja ili naredbi) ili Goffmanove (1959, 1967) **dramaturške analize** (koja promatra društvenu interakciju kao izvođenje uloga na pozornici), razvijene su s pretpostavkom o **ljudskim akterima** koji posjeduju svijest, namjeru (intencionalnost), emocije i sposobnost interpretacije društvenog konteksta iznutra. AI sustavi, koliko god bili sofisticirani, trenutno ne posjeduju ove atribute u ljudskom smislu. Njihovo ponašanje proizlazi iz kompleksnih

statističkih korelacija naučenih iz podataka, a ne iz unutarnjih mentalnih stanja ili svjesnih namjera. Ipak, sama činjenica da ljudi percipiraju AI agente kao društvene aktere, uspostavljuju s njima određene oblike odnosa i prilagođavaju svoja komunikacijska pravila u interakciji s njima, ukazuje na potrebu za **premošćivanjem konceptualnog jaza** između klasičnih komunikoloških i sociolingvističkih teorija i novih realnosti tehnološki posredovane komunikacije (Herring, 2020; Guzman, 2019). Potrebna je **nova sinteza** koja može objasniti kako se značenje i društveni kontekst konstruiraju u ovoj hibridnoj interakciji čovjek-stroj. Razumijevanje ove dinamike ključno je ne samo za analizu posljedica koje LLM-ovi imaju na postojeće društvene prakse, već i za promišljanje o novim mogućnostima koje otvaraju za buduće oblike suradnje, učenja, kreativnosti, pa čak i emocionalne podrške.

Primjer **personaliziranih edukativnih asistenata** dodatno ilustrira ovu ulogu LLM-ova kao graditelja konteksta. AI tutor koji radi s učenikom osnovne škole mora usvojiti potpuno drugačiji ton, rječnik i pedagoški pristup od onoga koji komunicira sa studentom na fakultetu ili stručnjakom na usavršavanju. Model koji se fino podešava za rad s mlađom djecom mora uskladiti točnost objašnjenja (npr. matematičkog koncepta) s korištenjem jednostavnog jezika, primjera iz dječjeg svijeta, te ohrabrujućeg i strpljivog tona. Mora znati kada postaviti pitanje da provjeri razumijevanje, kada ponuditi pomoći, a kada pustiti učenika da sam pokuša riješiti problem. Sve ovo zahtijeva implicitno ili eksplicitno poznavanje ne samo sadržaja koji se podučava, već i **jezičnih normi** ciljane dobne skupine, **razvojne psihologije** i **efikasnih pedagoških strategija** (Zou, Li & Yang, 2021; Khan Academy, 2023). Na taj način, LLM prelazi iz uloge pukog isporučitelja informacija u ulogu **aktivnog sudionika u oblikovanju obrazovnog konteksta i prakse**.

Postaje sve jasnije da veliki jezični modeli, utjelovljeni u komunikacijskim agentima, djeluju kao **dinamični suigrači u neprestanoj igri izgradnje i transformacije društvenog konteksta**. Njihova sposobnost da uče, prilagođavaju se i uvjerljivo simuliraju aspekte ljudske komunikacije otvara vrata potencijalno inkluzivnijoj, personaliziranoj i efikasnijoj interakciji, olakšavajući međukulturalno razumijevanje i nudeći nove oblike podrške i suradnje. Istovremeno, ta ista sposobnost nosi sa sobom rizike amplifikacije nejednakosti, širenja manipulativnih praksi i prenošenja netočnih ili pristranih obrazaca. Pomak u percepciji – od gledanja na LLM-ove kao na puki "alat" prema prepoznavanju njihove uloge kao **su-aktera** koji aktivno doprinose stvaranju značenja, normi i vrijednosti unutar zajednice – ključan je. Ova nova uloga, koja se tek počinje shvaćati u svojoj punoj kompleksnosti, zahtijeva razvoj snažnih etičkih standarda, regulatornih mehanizama i inovativnih teorijskih koncepata unutar komunikacijskih znanosti. Buduće generacije LLM-ova i AI agenata vjerojatno će imati još dublji utjecaj, aktivno "pregovarajući" našu percepciju stvarnosti i redefinirajući same temelje ljudske interakcije u digitalnom dobu.

Stoga, da bismo shvatili puni opseg njihove uloge, ključno je istražiti kako ovi algoritamski sugovornici **interveniraju u procesu oblikovanja individualnog identiteta, konstruiranja društvene stvarnosti, navigacije apstraktnim konceptima i facilitiranja kolektivnog djelovanja**.

#### 4.4.1 Jezik kao Ogledalo i Kalup Identiteta: Utjecaj AI Agenata na Naše "Ja"

Jezik funkcioniра i kao pozornica na kojoj se naši **identiteti** – osjećaj vlastitog "ja", pripadnosti grupama i pozicije u društvenoj strukturi – neprestano konstruiraju, izražavaju i pregovaraju. Način na koji govorimo, riječi koje biramo, stil koji usvajamo, pa čak i dijalekt ili naglasak koji koristimo, su markeri našeg osobnog, socijalnog i kulturnog identiteta (Gumperz, 1982;

Bucholtz & Hall, 2005). Komunicirajući s i preko **AI agenata** pokretanih velikim jezičnim modelima, ovaj intimni odnos između jezika i identiteta ulazi u novu, kompleksnu fazu. Interakcija s AI sustavima, koji generiraju i moderiraju sadržaj, te aktivno usvajaju i promoviraju određene jezične norme, pretvara jezični prostor u **hibridni okvir**. U tom okviru naš osjećaj sebe više nije oblikovan isključivo kroz dijalog s drugim ljudima, već i kroz stalnu, često nesvesnu, interakciju s algoritmima koji mogu suptilno (ili manje suptilno) utjecati na naš jezični izričaj, a time i na percepciju vlastitog identiteta (Herring, 2020; Hancock et al., 2020).

Utjecaj AI agenata na oblikovanje identiteta najočitiji je u **strukturiranim okruženjima**, poput poslovног svijeta ili obrazovnih institucija. Zamislimo, primjerice, **AI asistenta za ljudske resurse (HR)** koji pomaže zaposlenicima u snalaženju s internim procedurama, ispunjavanju obrazaca ili čak pripremi za godišnje razgovore. Agent će vjerojatno temeljem svojeg dizajna koristiti specifičan **profesionalni žargon**, održavati visok stupanj formalnosti i poticati korištenje preciznih, "poslovno prihvatljivih" izraza. Zaposlenik koji redovito komunicira s ovim agentom – tražeći informacije, dobivajući upute ili čak vježbajući odgovore – može nesvesno početi usvajati taj isti jezični registar i stil. Korištenje formalnijeg obraćanja, usvajanje novih stručnih termina i izbjegavanje neformalnih izraza postaje norma ne samo u komunikaciji s agentom, već potencijalno i u komunikaciji s ljudskim kolegama. Na taj način, AI agent djeluje kao **model i implicitni učitelj** korporativnog jezika, a zaposlenik, prilagođavajući svoj jezik, istovremeno prilagodava i svoj **profesionalni identitet**, signalizirajući kompetentnost i pripadnost organizacijskoj kulturi (Nass & Moon, 2000). Sličan proces dogada se u **obrazovnim postavkama**. Chatbot dizajniran za **učenje stranog jezika** osim što pomaže učeniku s gramatikom i vokabularom, izlaže ga i specifičnim **socio-kulturnim konvencijama** ciljane jezične zajednice – kako se ispravno pozdravlja, kako se izražava ljubaznost, koji su uobičajeni idiomi (Chaudhuri & Boonthum-Denecke, 2018). Povlačenje prema standardnom jeziku i kulturnim normama u jednu ruku izuzetno korisno za učenikovu integraciju i komunikacijsku kompetenciju. Međutim, istovremeno može stvoriti pritisak potiskivanja vlastitih **lokalnih dijalekata, regionalnih naglasaka ili osobnog, autentičnog načina izražavanja**, potencijalno vodeći ka osjećaju otuđenja od vlastitog jezičnog nasljeđa ili stvaranju uniformnijeg jezičnog identiteta.

Napetost između prilagodbe standardu i očuvanja jezične raznolikosti postaje još izraženija kada uzmemu u obzir korisnike iz **manjinskih jezičnih ili kulturnih zajednica**. Većina najnaprednijih LLM-ova i dalje je primarno trenirana na ogromnim korpusima engleskog jezika i podataka koji odražavaju dominantne zapadne perspektive. Kada korisnik čiji materinji jezik nije engleski, ili koji pripada kulturi s drugačijim komunikacijskim normama, ude u jezičnu interakciju s takvim agentom, može se suočiti sa sustavom koji ne razumije u potpunosti njegove jezične specifičnosti ili kulturni kontekst. Agent može preferirati ili bolje razumjeti standardni jezik, i na taj način posredno signalizirajući da su lokalni dijalekti, kreolski jezici ili specifični kulturni izrazi "manje vrijedni" ili "neispravni". U **online zajednicama**, posebice onima s velikim brojem korisnika gdje AI moderatori ili alati za prevođenje igraju značajnu ulogu, ovo može dovesti do postupne **homogenizacije jezika** – napuštanja lokalnih govornih oblika, žargona ili čak tradicionalnih metafora u korist globalno razumljivijeg, ali često i siromašnijeg, standardiziranog jezika (Blodgett et al., 2020; Mishra et al., 2022). Komunikacijski agenti, zamišljeni kao praktični i efikasni kanali, tako postaju nesvesni akteri u procesu **potiskivanja jezične raznolikosti** i erozije jedinstvenih jezičnih identiteta. Zamislimo tinejdžera iz određene regije koji koristi popularni chatbot; ako chatbot stalno ispravlja njegove lokalne izraze ili ne razumije regionalni humor, tinejdžer bi mogao početi izbjegavati te elemente svog identiteta u

online komunikaciji kako bi bio "bolje shvaćen" od strane AI-ja (i posredno, od šire online zajednice).

Paradoksalno, isti AI agenti mogu djelovati i kao moćni **mentor i čuvari specifičnih identiteta**, posebice unutar definiranih grupa ili organizacija. U sustavima za **upravljanje znanjem** (eng. **knowledge management**) unutar tvrtki, LLM-ovi ugradeni u interne tražilice ili asistente pomažu zaposlenicima da brzo pronađu relevantne informacije, razumiju složenu internu dokumentaciju ili čak dobiju savjete o najboljim praksama unutar tvrtke (Rae et al., 2021). Pritom, agent često usvaja specifičan **registar i stil** te organizacije. Komunicirajući s agentom, novi zaposlenik implicitno usvaja "način na koji se ovdje govori", percipirajući stil agenta kao autoritativen i poželjan. Ovo može ubrzati proces **organizacijske socijalizacije** i pomoći u oblikovanju identiteta zaposlenika kao kompetentnog člana tima koji "govori istim jezikom". Slično, u **online interesnim grupama**, AI moderatori ili chatbotovi mogu biti programirani da potiču korištenje specifičnog žargona ili referenci vezanih uz tu grupu, čime jačaju osjećaj zajedničkog identiteta i pripadnosti među članovima.

Čak i u izrazito osobnim domenama, poput **terapijskog konteksta**, AI agenti utječu na jezik i identitet. Chatbotovi dizajnirani za pružanje podrške kod anksioznosti ili depresije često koriste specifičan **terapijski diskurs**, temeljen na kognitivno-bihevioralnim ili drugim psihoterapijskim pristupima (npr. fokusiranje na misli, osjećaje i ponašanja, korištenje tehnika reframinga). Korisnik koji redovito komunicira s takvim agentom može pronaći olakšanje i korisne strategije (Fitzpatrick et al., 2017). Međutim, postoji i suptilni rizik da korisnik počne internalizirati taj specifični način govora o vlastitim iskustvima, potencijalno **suzvajući svoj autentični emocionalni vokabular** ili način izražavanja u korist "terapijski prihvatljivih" formulacija (Turkle, 2011). Dok strukturirani jezik može biti koristan za organiziranje misli i osjećaja, prekomjerna prilagodba unaprijed definiranom diskursu može dugoročno ograničiti individualnu ekspresivnost i jedinstvenost osobnog narativa.

Uloga komunikacijskih agenata, kako naglašavaju novija istraživanja, sve više prelazi u sferu aktivnog "**kuriranja" jezične i društvene dinamike** (Rosa & Flores, 2017; Weidinger et al., 2022). Ovisno o podacima na kojima su trenirani, načinu na koji su dizajnirani i ciljevima koje im developeri postave, AI agenti mogu postati **promotori jezične raznolikosti** – primjerice, ako su eksplicitno dizajnirani da podržavaju više jezika i dijalekata, ili da pomažu u očuvanju ugroženih jezika. Mogu njegovati **autentične lokalne govore** i pomagati u jačanju specifičnih **kulturnih identiteta**. Suprotno tome, ako su vođeni isključivo logikom globalnog tržišta i trenirani na dominantnim jezicima, mogu nesvesno (ili svjesno) postati **agenti jezične homogenizacije**, jačajući utjecaj globalnog standardnog jezika na štetu lokalnih varijeteta. Ova napetost između globalne povezanosti i lokalne autentičnosti postaje centralno pitanje u dizajnu i implementaciji AI komunikacijskih tehnologija.

Interakcija s LLM-ovima i komunikacijskim agentima, iako tek sliči na razmjenu informacija predstavlja dinamičan proces koji aktivno utječe na način na koji koristimo jezik, a time i na način na koji konstruiramo i izražavamo svoje **identitete**. Ovi AI akteri djeluju kao modeli, učitelji, čuvari normi i potencijalni homogenizatori jezičnog ponašanja. Njihov utjecaj zahtijeva dublje etičko, sociolingvističko i kognitivno promišljanje. Kako možemo dizajnirati AI sustave koji podržavaju bogatstvo ljudske jezične i identitetske raznolikosti, umjesto da je potiskuju? Kako osigurati da tehnološke inovacije ne vode ka uniformnoj, globalnoj komunikacijskoj monotoniji, već prema kreativnom širenju izražajnih mogućnosti uz poštovanje individualnih i kolektivnih identiteta? Pronalaženje odgovora na ova pitanja ključni je izazov za budućnost

komunikacije u doba umjetne inteligencije, jer način na koji ćemo oblikovati naše AI sugovornike neizbjježno će oblikovati i nas same.

#### 4.4.2 Jezik, AI Agenti i (Re)konstrukcija Stvarnosti: Algoritamske Leće i Okviri Značenja

Hipoteza o **lingvističkoj relativnosti**, najpoznatija kroz radeve Edwarda Sapira i Benjamina Lee Whorfa (Whorf, 1956; Lucy, 1992), a kasnije revitalizirana i nijansirana u suvremenim kognitivnim znanostima (Boroditsky, 2011; Gentner & Goldin-Meadow, 2003), sugerira da jezik aktivno **oblikuje i usmjerava način na koji percipiramo, kategoriziramo i razumijemo svijet oko nas**. Strukture našeg jezika – njegova gramatika, vokabular, metafore – mogu utjecati na to koje aspekte stvarnosti smatramo važnima, kako konceptualiziramo apstraktne ideje poput vremena ili prostora, pa čak i kako rezoniramo o uzročno-posljedičnim vezama. U eri velikih jezičnih modela, ova ideja dobiva novu, potencijalno znatno snažniju dimenziju. Komunikacijski agenti pokretani LLM-ovima postaju sveprisutni **posrednici u našem pristupu informacijama**, preuzimajući ulogu selekcioniranja, sažimanja, prevodenja, objašnjavanja i, u konačnici, **reprezentiranja stvarnosti** za nas. Time se otvara ključno pitanje: kako "kognitivna leća" samih jezičnih modela, oblikovana podacima na kojima su trenirani i algoritmima koji upravljaju njihovim odgovorima, utječe na našu vlastitu interpretaciju svijeta?

Recentna istraživanja i kritičke analize LLM-ova (Bender et al., 2021; Bommasani et al., 2021; Weidinger et al., 2022) snažno upozoravaju da stvarnost, promatrana kroz prizmu ovih modela, nije neutralna niti objektivna refleksija svijeta. Ona je nužno **filtrirana i oblikovana** prema statističkim obrascima, dominantnim narativima i inherentnim vrijednostima prisutnim u masivnim, ali neizbjegivo nesavršenim i pristranim korpusima podataka na kojima su modeli trenirani. Sam "sklop znanja" koji LLM posjeduje, iako naizgled enciklopedijski širok, nije rezultat kritičkog promišljanja ili izravnog iskustva, već **jezični konstrukt** – mozaik sastavljen od fragmenata tekstova i obrazaca izvučenih iz digitalnih arhiva. Stoga, kada AI agent odgovara na naše pitanje, sažima vijest ili objašnjava kompleksan koncept, on ne nudi nužno činjeničnu neutralnost, već **interpretaciju** oblikovanu podacima iz kojih crpi i algoritmima koji usmjeravaju proces generiranja (inferencije).

Moglo bi se reći da LLM-ovi, poput ljudskih jezičnih zajednica, grade vlastite **"okvire značenja"** (**framing**) (Lakoff, 2004; Entman, 1993). Oni uče koje riječi i koncepti se često pojavljuju zajedno, koje perspektive su dominantne u određenim diskursima, i koje aspekte problema treba naglasiti, a koje zanemariti. Na primjer, kada odgovara na pitanje o složenom društvenom problemu poput siromaštva, model čiji su podaci za treniranje pretežno sadržavali tekstove koji naglašavaju individualnu odgovornost mogao bi implicitno "uokviriti" problem na taj način, umanjujući važnost strukturnih faktora. Slično, model treniran primarno na zapadnim internetskim izvorima može imati poteškoća u adekvatnom predstavljanju ili razumijevanju narativa, koncepata ili vrijednosti specifičnih za nezapadne kulture, predstavljajući ih kroz zapadnocentričnu leću ili ih jednostavno marginalizirajući zbog slabije zastupljenosti u podacima (Kuhlmann et al., 2022; Prabhu & Birhane, 2021). Time se pluralnost ljudskih perspektiva može nenamjerno susiti, a dominantni pogledi na svijet dodatno učvrstiti. Zamislimo studenta koji koristi AI agenta za istraživanje povijesti kolonijalizma; ako agent primarno crpi iz izvora koji opravdavaju ili umanjuju negativne posljedice kolonijalizma, student može dobiti iskrivljenu i nepotpunu sliku tog kompleksnog povijesnog razdoblja.

Nadalje, zbog svoje statističke prirode, LLM-ovi mogu nehotice **amplificirati** one teme, mišljenja ili izvore informacija koje prepoznaju kao **najistaknutije ili najčešće** u podacima za treniranje, pridajući im neproporcionalnu važnost. Ovo može dovesti do situacije gdje korisnik, tražeći informacije o nekoj temi, dobiva odgovore koji pretežno odražavaju mainstream ili popularna stajališta, dok su alternativne, manje zastupljene ili kontroverzne perspektive **selektivno marginalizirane** ili potpuno izostavljene (Zhao et al., 2021; Pariser,

2011 - koncept "filter bubble"). Korisnik tako, umjesto uravnoteženog pregleda različitih gledišta, može dobiti iskrivljenu ili manjkavu sliku stvarnosti, oblikovanu algoritamskom preferencijom prema popularnosti ili učestalosti. U tim slučajevima, jezik modela doslovno funkcioniра kao **kognitivna leća**: Al aktivno **konstruira interpretaciju stvarnosti** za korisnika. Ako korisniku nedostaje kritička distanca ili svijest o načinu funkcioniranja modela, postoji realan rizik da će nesvesno usvojiti te algoritamski generirane interpretativne okvire kao objektivnu istinu.

Naravno, **interakcija s AI agentima i njihov utjecaj na interpretaciju stvarnosti** nisu jednoznačni i ovise o samim korisnicima. Neki su korisnici skloniji automatskom **internaliziranju i reproduciraju** jezičnih struktura, fraza i interpretativnih okvira koje im model nudi, posebice ako agenta percipiraju kao autoritativen izvor znanja. Drugi, posebice oni s razvijenjom **metalingvističkom svjesnošću** ili kritičkim mišljenjem, mogu koristiti interakciju s agentom kao priliku za **propitivanje vlastitih prepostavki i re-konceptualizaciju** problema (Shin et al., 2020). Oni mogu postavljati agentu izazovna pitanja, tražiti alternativne perspektive ili uspoređivati njegove odgovore s drugim izvorima informacija, koristeći AI kao "isporučitelja" dodatnih gledišta, a ne kao konačnog "tumača istine". Ipak, u širem društvenom kontekstu, postoji opravdana zabrinutost da bi nekritičko i masovno oslanjanje na interpretativne obrasce koje nude LLM-ovi moglo dovesti do **ujednačavanja mišljenja, simplifikacije složenih pitanja** i općenito **osromašenja javnog diskursa**.

Jedan od najočitijih mehanizama kroz koje AI agenci mogu iskriviti percepciju stvarnosti jesu **pristranosti** naslijedene iz korpusa za treniranje. Ako model uči iz tekstova koji perpetuiraju rodne stereotipe (npr. povezujući žene primarno s kućanskim poslovima, a muškarce s tehnologijom ili vodstvom), političke predrasude (npr. negativno prikazujući određene političke skupine) ili druge oblike netočnih i štetnih generalizacija, on će te obrasce vjerojatno reproducirati u svojim odgovorima (Lucy & Bamman, 2021; Bolukbasi et al., 2016). Ove algoritamski generirane pristranosti ne samo da mogu iskriviti korisnikov doživljaj stvarnosti (npr. navodeći ga da vjeruje u neistinite stereotipe), već mogu suptilno, ali postojano **oblikovati njegove stavove, uvjerenja i odluke**, vodeći ga prema određenim narativima na štetu drugih. Zamislimo AI agenta koji pruža vijesti; ako je pristran prema određenoj političkoj strani, njegovi sažeci vijesti ili odgovori na pitanja o aktualnim događajima mogu suptilno favorizirati tu stranu, utječući na političke stavove korisnika. Uloga AI agentovih odgovora time prelazi granicu neutralnog informiranja i ulazi u sferu aktivnog **oblikovanja mišljenja**, što zahtijeva posebnu pažnju i kontrolu interpretativnih okvira, posebice u osjetljivim domenama poput obrazovanja, medija i formiranja javnog mnenja.

Suočeni s ovim izazovima, istraživači i developeri aktivno istražuju različite tehnike "**kontrole**" ili "**ispravljanja**" interpretativnih okvira koje LLM-ovi nameću. To uključuje već spomenute napore u **pažljivijem odabiru i filtriranju podataka** za treniranje, **razvoj algoritamskih tehnika za debiasing** (Ribeiro et al., 2020; Dugan et al., 2022), te **poboljšanje metoda poravnjanja** kako bi se modeli usmjerili prema generiranju uravnoteženijih, činjenično točnijih i manje pristranih odgovora. Neki pristupi pokušavaju eksplicitno povećati **kulturnu raznolikost perspektiva** u modelima, bilo kroz uključivanje raznolikijih podataka ili kroz specifične tehnike finog podešavanja. Međutim, ove strategije nisu samo tehnički izazovi; one neizbjegivo uključuju i složene **sociolingvističke i političke rasprave**: Čija se gledišta smatraju "vjerodostojnjima" ili "poželjnima"? Kako definirati i mjeriti "točnost", "neutralnost" ili "uravnoteženost" u kontekstu inherentno spornih društvenih pitanja? Koji su kriteriji za postizanje istinski pluralnog razumijevanja složenih fenomena? Samo priznavanje da LLM-ovi nisu neutralni promatrači, već

aktivni sudionici u oblikovanju "zajedničke" društvene stvarnosti, prvi je korak prema odgovornijem pristupu.

U praksi, put prema ublažavanju negativnih utjecaja LLM-ova na interpretaciju stvarnosti vjerojatno leži u kombinaciji pristupa. **Regulacija** može igrati ulogu u zahtijevanju veće **transparentnosti** u pogledu podataka za treniranje i algoritama, kao i u postavljanju standarda za točnost i pravednost u određenim visoko rizičnim primjenama. Istovremeno, jednostavna **cenzura** ili ograničavanje sadržaja nailazi na legitimne kritike vezane uz slobodu govora i potencijalno može prikriti, umjesto riješiti, temeljne probleme pristranosti. Stoga se naglasak sve više stavlja na **promicanje raznolikosti izvora podataka**, aktivno **uključivanje perspektiva iz nedovoljno zastupljenih zajednica** u proces razvoja i evaluacije, te, što je možda najvažnije, na **sustavno obrazovanje korisnika** o prirodi LLM tehnologije.

Razvijanje **kritičke AI pismenosti** – sposobnosti razumijevanja kako ovi modeli rade, prepoznavanja njihovih potencijalnih pristranosti i ograničenja, te vještine kritičkog vrednovanja informacija koje generiraju – ključno je za osnaživanje korisnika. Umjesto da pasivno prihvaćaju odgovore AI agenata kao konačnu istinu, informirani korisnici mogu ih koristiti kao **polazišnu točku za istraživanje**, kao **izvor dodatnih perspektiva**, ili kao **alat za propitivanje vlastitih prepostavki**, zadržavajući pritom agensnost u vlastitom procesu formiranja mišljenja i interpretacije svijeta.

Zaključno, razgovorni agenti i veliki jezični modeli utjelovljuju **dvojaku moć** u odnosu na našu percepciju stvarnosti. S jedne strane, nude neviđenu efikasnost u pristupu informacijama, potencijal za premošćivanje jezičnih i kulturnih barijera, te nove načine za učenje i razumijevanje kompleksnih tema. S druge strane, njihovi suptilni, algoritamski vođeni konceptualni odabiri, pristranosti naslijedene iz podataka i inherentna ograničenja u razumijevanju mogu neprimjetno, ali snažno **usmjeravati naš pogled na svijet**, potencijalno vodeći ka homogenizaciji mišljenja, širenju dezinformacija i iskrivljenoj percepciji stvarnosti. Ključni izazov za budućnost leži u pronalaženju načina da se ova moćna tehnologija iskoristi na način koji **promiče pluralnost mišljenja, kritičko promišljanje i dublje razumijevanje**, umjesto da nas vodi prema pojednostavljenoj, algoritamski kuriranoj i potencijalno manipuliranoj verziji svijeta. To zahtijeva ne samo tehničke inovacije, već i duboku etičku refleksiju i društveni angažman u oblikovanju budućnosti komunikacije u doba umjetne inteligencije.

#### **4.4.3 Jezik kao most ka apstrakciji: Stvaranje i dijeljenje nevidljivih svjetova**

Jedna od najdubljih i najtransformativnijih moći ljudskog jezika leži u njegovoj sposobnosti da nas odvoji od okova neposredne, osjetilne stvarnosti i otvori vrata beskrajnim prostranstvima **apstraktног mišljenja**. Dok mnoge životinske vrste posjeduju sofisticirane sustave komunikacije za signaliziranje opasnosti, pronalaženje hrane ili koordinaciju grupe unutar *ovdje i sada*, ljudski jezik posjeduje jedinstvenu sposobnost da kodira, manipulira, dijeli i transformira **ideje koje nemaju direktni fizički korelat**. Bilo da se radi o konceptima poput pravde, slobode ili ljubavi, o zamišljanju događaja iz daleke prošlosti ili planiranju kompleksnih scenarija za hipotetsku budućnost, o formuliranju znanstvenih zakona koji opisuju nevidljive sile, ili o stvaranju svjetova maště kroz književnost i umjetnost – jezik djeluje kao fundamentalni mehanizam koji omogućuje ove skokove izvan konkretnog. On je **alat za konstrukciju i navigaciju kroz nevidljive svjetove osjećaja, misli, vrijednosti i mogućnosti**, ključan za naš kognitivni razvoj, kulturnu evoluciju i intelektualni napredak (Vygotsky, 1986; Deacon, 1997; Pinker, 1994).

Ovaj potencijal jezika za apstrakciju bio je pokretačka snaga ljudske civilizacije od samih početaka. Kroz **usmenu predaju, priče, mitove i legende**, rane ljudske zajednice nisu samo prenosile praktična znanja, već su gradile složene simboličke sustave koji su objašnjavali porijeklo svijeta i čovjeka, uspostavljali moralne i etičke kodekse, definirali društvene uloge i njegovali osjećaj kolektivnog identiteta i svrhe. Mitovi o stvaranju, priče o bogovima i herojima, legende o precima – sve su to jezične konstrukcije koje su omogućavale ljudima da pronađu smisao u prirodnim pojavama (poput izmjene godišnjih doba ili nebeskih kretanja), da se nose s egzistencijalnim pitanjima (poput života, smrti i patnje) i da uspostave zajedničke vrijednosti koje su regulirale društveni život (Campbell, 1949; Eliade, 1963). Moć narativa – sposobnost jezika da isplete slijed događaja, uvede likove s motivacijama i stvari emocionalni luk – pokazala se kao izuzetno efikasan način za prijenos kompleksnih ideja i vrijednosti kroz generacije, mnogo prije pojave formalnog obrazovanja ili znanstvenog diskursa.

Lav Vigotski (Vygotsky, 1986), u svojim utjecajnim radovima o odnosu jezika i misli, naglasio je presudnu ulogu jezika u **kognitivnom razvoju pojedinca**. Tvrđio je da usvajanje jezika, posebice kroz socijalnu interakciju, omogućuje djetetu da prijede s konkretnog, situacijski vezanog razmišljanja na više, **apstraktne oblike mišljenja**. Jezik pruža "psihološke alate" – riječi i koncepte – koji omogućuju internalizaciju znanja, planiranje djelovanja neovisno o neposrednim podražajima, te razvoj sposobnosti za logičko zaključivanje, samorefleksiju i svjesnu kontrolu vlastitih kognitivnih procesa (kroz "unutarnji govor"). Usvajanjem jezika, dijete osim učenja imenovanja stvari oko sebe, usvaja i kulturno oblikovane načine kategorizacije svijeta i razmišljanja o njemu.

Posebno značajna manifestacija apstraktne moći jezika jest sposobnost **mentalnog putovanja kroz vrijeme**. Za razliku od drugih vrsta čija je svijest uglavnom vezana uz sadašnji trenutak, ljudi mogu jezikom detaljno opisivati **događaje iz prošlosti** – sjećanja, povijesne narative, naučene lekcije. Ovo "premještanje u prošlost" (displacement) ključno je za učenje iz iskustva (vlastitog i tuđeg), za izgradnju osobnog i kolektivnog pamćenja, te za osiguravanje kontinuiteta znanja i kulturnih vrijednosti. Jednako tako, jezik nam omogućuje da **zamišljamo, planiramo i komuniciramo o budućnosti**. Možemo postavljati dugoročne ciljeve, razvijati strategije, predviđati posljedice različitih akcija i koordinirati zajedničke napore prema željenim ishodima. Ova sposobnost projekcije u budućnost, utemeljena na jezičnoj manipulaciji vremenom i mogućnostima, u korijenu je ljudskog planiranja, inovacija i društvenog napretka, od poljoprivredne revolucije do slanja ljudi na Mjesec.

Nezaobilazna je i uloga jezika u artikuliranju **hipotetičkih scenarija** i istraživanju svijeta "što ako?". Kroz gramatičke strukture poput kondicionala ("ako...onda...") i subjunktiva, te kroz bogatstvo metafora i analogija, jezik nam dopušta da kreativno prekoračimo granice postojeće stvarnosti i istražujemo alternativne mogućnosti. U znanosti, ova sposobnost je fundamentalna. Znanstvene teorije često započinju kao **hipotetički modeli** koji pokušavaju objasniti promatrane fenomene; **misaoni eksperimenti** (poput Einsteinovog zamišljanja vožnje na zraci svjetlosti ili Schrödingerove mačke) omogućuju istraživanje posljedica teorija u ekstremnim ili nemogućim uvjetima; a **matematički jezik** pruža formalni okvir za precizno modeliranje i predviđanje ishoda kompleksnih procesa, od kretanja planeta do širenja epidemija.

Naravno, **umjetničko izražavanje** predstavlja vrhunac jezične sposobnosti za prenošenje apstraktnih ideja i iskustava. Književnost, poezija, drama – sve se one oslanjaju na moć jezika da evocira složene emocije, stvori živopisne mentalne slike, istraži duboke filozofske teme i prenese simbolička značenja koja često nadilaze doslovnu interpretaciju (Eagleton, 1996). Metafore, aluzije, ritam, zvuk – sve su to jezični alati kojima umjetnici oblikuju našu percepciju i razumijevanje ljudskog stanja. Čak i drevni **mitovi i arhetipovi** (Jung, 1968; Campbell, 1949) nastavljaju živjeti i rezonirati u modernoj popularnoj kulturi. Priče o superherojima koji se bore protiv zla, romantične komedije koje istražuju potragu za ljubavlju, znanstveno-fantastični epovi koji propituju budućnost čovječanstva – sve su to suvremene manifestacije prastare ljudske potrebe da kroz narativ istražujemo univerzalne teme, vrijednosti i strahove, koristeći jezik kao medij za kolektivno sanjarenje i promišljanje.

Važno je također i prepoznati da je sposobnost jezika za apstrakciju neodvojiva od njegove **dinamičnosti i inovativnosti**. Jezik nije statican sustav; on se neprestano mijenja i prilagođava kako bi odgovorio na nove potrebe, tehnologije i kulturne trendove. Svako doba stvara **nove riječi (neologizme)** i **nove izraze** za opisivanje novih pojava – od pojnova vezanih uz industrijsku revoluciju, preko rječnika psihoanalize, do današnjeg vokabulara interneta ("googlati", "tweetati", "selfie", "algoritam") i umjetne inteligencije ("prompt", "halucinacija", "agent"). **Znanstvena terminologija** posebno je dobar primjer ove precizne jezične inovacije; pojmovi poput "kvant", "crna rupa", "DNK" ili "neuron" nisu samo oznake, već konceptualni alati koji omogućuju znanstvenicima da precizno komuniciraju, grade složene teorije i dalje istražuju temeljne zakone prirode (Coulmas, 2003). Ova stalna evolucija jezika osigurava da on ostaje relevantan i moćan alat za artikuliranje sve kompleksnijih apstraktnih ideja koje oblikuju naš svijet.

U kontekstu ove inherentne moći jezika za apstrakciju, pojava **velikih jezičnih modela (LLM-ova)** predstavlja značajan novi faktor. Istrenirani na nezamislivim količinama ljudskog teksta koji eksplicitno i implicitno sadrži svu tu akumuliranu apstraktну misao, AI sustavi su moći **akteri u procesu dijeljenja, objašnjavanja, pa čak i generiranja apstraktnih ideja**. Njihov utjecaj očituje se na više načina. Pripe svega, veliki jezični modeli olakšavaju pristup kompleksnim konceptima široj populaciji, što se ogleda u njihovoj mogućnosti pojednostavljanja filozofskih ili znanstvenih ideja. Tako korisnik bez formalnog filozofskog obrazovanja može dobiti jasno objašnjenje Kantovog kategoričkog imperativa kroz praktične primjere iz svakodnevnog života. Student fizike, s druge strane, može dobiti intuitivno objašnjenje Einsteinove teorije relativnosti, prilagođeno njegovom predznanju (Brown et al., 2020; Gao et al., 2021). Sve se to može predstaviti u interaktivnom kontekstu, uporabom personaliziranih objašnjenja, primjere i vježbe, Chi et al., 2021; Khan Academy, 2023).

Osim demokratizacije znanja, veliki jezični modeli igraju važnu ulogu u poticanju kreativnosti i inovacija. Primjerice, kao partneri u kreativnom procesu, predlažu alternativne narativne

strukture ili scenarije razvoja priče. Na sličan način, znanstvenici koriste jezične modele za istraživanje novih hipoteza ili pristupa postojećim problemima. Kombinirajući informacije na neočekivane načine, LLM-ovi potiču ljudе na kreativno razmišljanje izvan standardnih paradigma i ustaljenih obrazaca (Google DeepMind, 2023b).

LLM-ovi također olakšavaju medukturnu razmjenu ideja kroz napredne sustave strojnog prevođenja koji nisu ograničeni na puko prenošenje riječi, već učinkovito prenose složene, apstraktne ideje unatoč jezičnih i kulturnih barijera. (Fan et al., 2021; Costa-jussà et al., 2022).

Konačno, modeli pokazuju rastuću sposobnost generiranja tekstova koji zahtijevaju duboko razumijevanje apstraktnih struktura, poput detaljnih analiza, argumentiranih eseja ili filozofskih dijaloga. Tu, čini se prednjače pristupi poput „lanca misli“ koji sustavima omogućuju kompleksne procese rezoniranja i generiranja koherentnih zaključaka, premdа je pitanje njihovog stvarnog „razumijevanja“ još uvjek predmet znanstvenih debata (Floridi & Chiriatti, 2020; OpenAI, 2024c). Usprkos tome, potencijal velikih jezičnih modela za uključivanje u složene diskurzivne tokove nesumnjivo je impresivan i važan za daljnje istraživanje njihove uloge u oblikovanju društvenih interakcija i apstraktog razmišljanja.

Nove komunikacijske prakse, međutim, dolaze s važnim **etičkim i filozofskim implikacijama**. Sposobnost LLM-ova da "simuliraju" rasprave o moralnim dilemama, prezentiraju različite etičke argumente ili čak analiziraju potencijalne posljedice odluka (Bai et al., 2022; Shin et al., 2020) postavlja pitanja o **izvoru i autoritetu** takvih "automatiziranih" moralnih promišljanja. Postoji li rizik da korisnici nekritički prihvate etičke okvire koje generira AI? Kako osigurati **točnost, pravednost i odgovornost** kada se AI koristi kao podrška u donošenju odluka s moralnim implikacijama? Granice između ljudskog i "automatiziranog" razmišljanja postaju sve zamagljivije, zahtijevajući nova promišljanja o ulozi tehnologije u oblikovanju naših najdubljih uvjerenja i vrijednosti.

U konačnici, ovo nas suočava s fundamentalnim pitanjima: Kako informacijsko društvo i umjetna inteligencija redefiniraju naše vlastite kognitivne sposobnosti? Koliko smo spremni prihvativati "automatizirano" generirane apstraktne ideje kao vrijedan doprinos ljudskoj misli? I, najvažnije, kako možemo osigurati da zadržimo kritičku kontrolu i ljudsku agensnost u ovom neprestanom procesu jezične i spoznajne razmjene, koristeći AI kao alat za proširenje naših horizontata, a ne kao zamjenu za vlastito promišljanje? Odgovori na ova pitanja oblikovat će ne samo budućnost tehnologije, već i budućnost ljudske misli same.

#### 4.4.4 Jezik i Kolektivno Stvaranje: Od Društvenih Ugovora do AI Suradnje

Jezik, u svojoj najdubljoj i najmoćnijoj funkciji, djeluje kao temeljni **operativni sustav za kolektivno stvaranje**. On je onaj složeni, neprekidni proces kroz koji ljudske zajednice grade, održavaju, pregovaraju i transformiraju zajedničke stvarnosti koje daleko nadilaze kapacitete bilo kojeg pojedinca. Naša jedinstvena sposobnost da jezikom artikuliramo zajedničke ciljeve, od izgradnje piramide do slanja rovera na Mars, da pregovaramo i kodificiramo pravila suživota, od nepisanih normi do složenih pravnih sustava, da akumuliramo i prenosimo znanje kroz generacije putem obrazovanja, znanosti i kulture, te da koordiniramo kompleksne akcije na velikoj skali u politici, ekonomiji ili umjetnosti – sve to počiva na postojanju zajedničkog jezičnog okvira. Jezik djeluje kao svojevrsno društveno ljestivo, omogućujući nam da premostimo individualne perspektive i stvorimo ono što filozofi poput Johna Searlea (1995) nazivaju "institucionalnim činjenicama" ili povjesničari poput Yuvala Hararija (2014) "zajedničkim imaginacijama". Koncepti poput novca, koji ima vrijednost samo zato što kolektivno vjerujemo da je ima, nacija, koje postoje kao zamišljene zajednice povezane zajedničkim narativima,

korporacija, kao pravnih fikcija s pravima i obvezama, ili apstraktnih idealnih ljudskih prava – sve su to moćne društvene konstrukcije koje duguju svoje postojanje i djelotvornost našoj sposobnosti da ih definiramo, komuniciramo i održavamo kroz jezik.

U ovaj već složeni ekosustav kolektivnog stvaranja, suvremeno doba uvodi novog, izuzetno moćnog aktera: **velike jezične modele (LLM-ove)**. Njihova rapidna integracija u digitalne alate i komunikacijske platforme predstavlja više od pukog tehnološkog unapređenja; ona aktivno intervenira i preoblikuje same procese kroz koje surađujemo, inoviramo i gradimo zajedničko znanje. LLM-ovi, sa svojom sposobnošću razumijevanja, generiranja, sažimanja, prevodenja i čak rezoniranja o tekstu na gotovo ljudskoj razini, prestaju biti samo pasivni izvori informacija. Njihova sposobnost da obrađuju i generiraju jezik čini ih **aktivnim sudionicima i potencijalnim katalizatorima** u našim suradničkim naporima. Oni posjeduju potencijal da značajno ubrzaju kolektivne procese, demokratiziraju pristup znanju potrebnom za sudjelovanje, ali istovremeno nose i rizike homogenizacije kreativnih ishoda ili čak potkopavanja temelja povjerenja i zajedničkog razumijevanja na kojima počiva svako kolektivno djelovanje.

Jedno od najočitijih područja gdje LLM-ovi transformiraju kolektivno stvaranje jest **poboljšanje kolaborativnog rada**, posebice u domenama koje se intenzivno oslanjaju na tekstualnu komunikaciju i produkciju. Zamislimo međunarodni istraživački tim koji radi na kompleksnom znanstvenom radu. Integrirani AI asistenti, poput onih ugrađenih u platforme kao što je **Microsoft 365 Copilot** ili specijaliziranim alatima za akademsko pisanje, mogu djelovati kao neumorni suradnici. Mogu generirati početne nacrte literaturnog pregleda na temelju ključnih riječi, pomoći u strukturiranju argumentacije, predlagati relevantne reference, osiguravati konzistentnost terminologije kroz cijeli dokument, pa čak i prevoditi dijelove teksta kako bi članovi tima s različitim materinjim jezicima mogli lakše suradivati (Brynjolfsson et al., 2023; Dowling & Zaki, 2023). Slično, u poslovnom svijetu, marketinški tim može koristiti LLM za brainstorming sloganata, generiranje različitih verzija promotivnog teksta prilagođenih različitim ciljnim skupinama, ili za brzo sažimanje i analizu sentimenta u povratnim informacijama od klijenata prikupljenim s društvenih mreža. Automatizacijom ovih često mukotrpnih aspekata pisanja, uređivanja i sinteze informacija, LLM-ovi oslobođaju dragocjeno vrijeme i kognitivne resurse ljudskih suradnika, omogućujući im da se usredotoče na više razine promišljanja – strateško planiranje, kritičku analizu, originalno rješavanje problema i kreativno usmjeravanje projekta. Međutim, ova nova dinamika čovjek-AI suradnje nije bez izazova. Postavljaju se nova pitanja: Kako osigurati da konačni proizvod zadrži koherentan i autentičan "glas", a ne postane generička mješavina ljudskog i strojnog stila? Kako efikasno upravljati rizikom da se AI-generirane netočnosti, "halucinacije" ili pristranosti neprimjetno uvuku u kolektivni rad i tamo multipliciraju (Weidinger et al., 2022)? Kako se mijenja osjećaj vlasništva i odgovornosti za zajednički rad kada AI postane značajan kontributor? Kako osigurati da AI alati ne zamijene, već nadopune ljudsku ekspertizu i kritičko prosudjivanje?

Nadalje, LLM-ovi posjeduju značajan potencijal za **demokratizaciju pristupa znanju i informacijama**, što je temeljni preduvjet za informirano građanstvo, kolektivno odlučivanje i poticanje inovacija iz svih dijelova društva. Njihova sposobnost da kompleksne tehničke, znanstvene, pravne ili filozofske koncepte "prevedu" na jednostavniji, razumljiviji jezik, prilagođen specifičnoj publici ili razini predznanja korisnika, može srušiti barijere koje su tradicionalno odvajale stručnjake od laika (OpenAI, 2023). Zamislimo lokalnu zajednicu koja raspravlja o prijedlogu izgradnje industrijskog postrojenja; LLM bi mogao pomoći građanima da razumiju tehničke detalje studije o utjecaju na okoliš, da protumače složene pravne implikacije prostornog plana, ili da analiziraju ekonomski argumente za i protiv, čak i ako nemaju

prethodno stručno znanje u tim područjima. Ovo ih osnažuje da formiraju informiranija mišljenja i aktivnije sudjeluju u demokratskom procesu. U globalnom kontekstu, znanstvenici ili studenti iz zemalja s manje resursa mogu koristiti LLM-ove za pristup i razumijevanje najnovijih istraživanja objavljenih u skupim časopisima ili na jezicima koje ne govore tečno, čime se potencijalno smanjuje globalna nejednakost u pristupu znanstvenim spoznajama. Ipak, obećanje demokratizacije dolazi s važnim ogradama. Postoji realan rizik da korisnici nekritički prihvate pojednostavljena ili čak pogrešna i pristrana objašnjenja koja generira AI, što može dovesti do iluzije razumijevanja ili širenja dezinformacija pod krinkom pristupačnosti (Bender et al., 2021). Također, sam pristup najnaprednjim LLM alatima i potrebnoj infrastrukturi (brzi internet, odgovarajući uređaji) još uvijek nije univerzalan, što prijeti stvaranjem nove **digitalne podjele** – jaza između onih koji mogu iskoristiti prednosti AI-poboljšanog pristupa znanju i onih koji ostaju isključeni (Vinuesa et al., 2020). Istinska demokratizacija stoga zahtjeva ne samo tehnološku dostupnost, već i široko rasprostranjeno obrazovanje o **kritičkoj AI pismenosti** – sposobnosti korisnika da odgovorno koriste, propituju, vrednuju pouzdanost, potencijalne pristranosti i ograničenja informacija dobivenih od AI sustava (Ng et al., 2023).

U sferi **umjetnosti i kreativnog izražavanja**, LLM-ovi otvaraju fascinantne, premda i kontroverzne, nove avenirje za kolektivno stvaranje koje uključuje ne-ljudske entitete. Umjetnici sve više eksperimentiraju s LLM-ovima i srodnim generativnim modelima (poput modela za generiranje slika kao što su Midjourney, Stable Diffusion ili DALL-E 3, te glazbenih modela poput Suno ili Udio) ne samo kao alatima, već kao **kreativnim suradnicima**, izvorima neočekivane inspiracije ili čak kao samostalnim "umjetnicima" (u određenom smislu) (Manovich, 2023; Companion et al., 2023; Mazzone & Elgammal, 2019). Proces često uključuje iterativni dijalog: umjetnik daje početnu ideju ili stilsku uputu (prompt), AI generira niz mogućnosti, umjetnik reagira, modificira prompt, odabire ili kombinira elemente, AI generira nove verzije, i tako dalje. Rezultat su **hibridna djela** – pjesme napisane u suradnji s AI-jem, vizualne kompozicije koje spajaju ljudsku estetiku s algoritamskom nepredvidljivošću, glazbene skladbe gdje AI generira harmonije ili ritmove koje ljudski glazbenik zatim interpretira i aranžira. Ova čovjek-AI ko-kreacija potiče eksperimentiranje s novim formama i stilovima, pomiče granice tradicionalnih umjetničkih disciplina i demokratizira pristup kreativnim alatima (omogućujući i ljudima bez formalne umjetničke obuke da vizualiziraju ili izraze svoje ideje). Istovremeno, ona pokreće fundamentalna filozofska i pravna pitanja o prirodi **autorstva** (čije je djelo ako je AI značajno doprinio?), **originalnosti** (je li AI-generirani sadržaj derivativan ili istinski nov?), **umjetničkoj vještini** (umanjuje li korištenje AI-ja vrijednost ljudskog truda i talenta?) i samoj **definiciji kreativnosti** – može li sustav koji nema svijest, emocije ili namjeru zaista biti "kreativan"? (Cheng et al., 2023; Epstein et al., 2023).

**Poticanje međukulturene komunikacije i suradnje** predstavlja još jedno područje gdje LLM-ovi mogu značajno doprinijeti kolektivnim naporima na globalnoj razini. Napredne sposobnosti strojnog prevodenja, posebice sposobnost novijih modela da bolje razumiju kontekst, idiome, kulturne reference i nijanse poput formalnosti (kao što pokazuju mogućnosti GPT-4o (OpenAI, 2024b)), ruše jezične barijere koje su tradicionalno otežavale suradnju između timova, organizacija i pojedinaca iz različitih dijelova svijeta (Costa-jussà et al., 2022; Fan et al., 2021). Zamislimo globalni tim znanstvenika koji radi na rješavanju klimatskih promjena; LLM-pokretani alati mogu omogućiti besprijecknu razmjenu ideja i podataka u stvarnom vremenu, prevodenje istraživačkih radova i vođenje sastanaka na više jezika istovremeno. Potencijal ide i dalje: LLM-ovi bi mogli djelovati kao "**kulturni medijatori**", suptilno pomažući korisnicima da prepoznaju i premoste razlike u komunikacijskim stilovima (npr. direktnost vs. indirektnost, niska vs. visoka kontekstualnost) koje često dovode do nesporazuma u multikulturalnim okruženjima. Na

primjer, agent bi mogao predložiti preoblikovanje e-pošte kako bi bila prikladnija za primatelja iz drugačije kulture. Ipak, kao što je ranije naglašeno, rizici **pristranosti i kulturne neosjetljivosti** su značajni. Ako modeli nisu adekvatno trenirani na raznolikim jezicima i kulturnim podacima, njihovi prijevodi mogu biti netočni, mogu propustiti ključne nijanse ili čak nametnuti dominantnu kulturnu perspektivu (Prabhakaran et al., 2022). Postizanje istinski pravedne i inkluzivne globalne komunikacije potpomognute AI-jem zahtijeva kontinuirane napore u razvoju višejezičnih i multikulturalnih modela te svijest o potencijalnim zamkama algoritamskog prevodenja kulture.

Unatoč svim ovim obećavajućim mogućnostima, važno je trezveno sagledati i značajne **izazove i etičke implikacije** koje korištenje LLM-ova unosi u procese kolektivnog stvaranja. Jedna od najozbiljnijih prijetnji jest potencijal za **masovnu proizvodnju i distribuciju dezinformacija, propagande i manipulativnog sadržaja**. Sposobnost LLM-ova da generiraju velike količine uvjerljivog, gramatički ispravnog i naizgled autorativnog teksta o bilo kojoj temi može se lako zloupotrijebiti za kreiranje lažnih vijesti, širenje teorija zavjere, automatizirano generiranje komentara na društvenim mrežama (astroturfing), ili za provođenje sofisticiranih, personaliziranih kampanja utjecaja na javno mnjenje na skali i brzini koje su prije bile nezamislive (Buchanan et al., 2021; Goldstein et al., 2023). Ovo predstavlja fundamentalnu prijetnju informiranom javnom diskursu, demokratskim procesima i samom povjerenju u informacije, jer postaje sve teže razlikovati autentičan sadržaj od sintetičkog.

Nadalje, postoji realan rizik od **homogenizacije ideja i stilova**. Ako se pojedinci i organizacije previše oslanjaju na iste, dominantne LLM alate za generiranje ideja, pisanje tekstova ili rješavanje problema, to može dovesti do smanjenja raznolikosti u pristupima, perspektivama i načinima izražavanja. Možemo završiti u svijetu gdje znanstveni radovi, marketinški materijali ili čak umjetnička djela počinju zvučati ili izgledati slično, obliskovani prema preferencijama i ograničenjima nekolicine dominantnih AI modela, čime se osiromašuje kolektivni intelektualni i kulturni krajolik (Liang et al., 2021). Pojačavanje postojećih **pristranosti** kroz AI alate također može imati razorne posljedice na kolektivne ishode – zamislimo AI sustav korišten u urbanističkom planiranju koji, zbog pristranih podataka, sustavno zanemaruje potrebe siromašnijih četvrti, ili AI alat za recenziranje znanstvenih radova koji favorizira određene metodologije ili istraživače iz prestižnih institucija (Mehrabi et al., 2021).

Konačno, pitanje **odgovornosti** postaje izuzetno složeno. Kada kolektivni projekt, u kojem je AI igrao značajnu ulogu, pode po zlu – bilo da se radi o finansijskom gubitku, štetnom utjecaju na okoliš, ili društvenoj nepravdi – kako utvrditi tko snosi odgovornost? Je li to ljudski tim, developeri AI alata, ili sam AI (što je pravno i filozofski problematično)? Neprozirnost ("problem crne kutije") funkcioniranja mnogih LLM-ova dodatno otežava praćenje uzročno-posljedičnih veza i dodjeljivanje odgovornosti (Burrell, 2016; Mittelstadt et al., 2019).

Konkretni primjeri iz prakse jasno ilustriraju ovu dvojaku prirodu utjecaja LLM-ova na kolektivno stvaranje. U znanstvenoj suradnji, alati poput Google DeepMindovog **FunSearch** (koji koristi LLM za otkrivanje novih matematičkih rješenja (Google DeepMind, 2023b)) ili sustava poput **AlphaFold** koji pomažu u analizi složenih bioloških podataka (Jumper et al., 2021) pokazuju kako AI može ubrzati otkrića i potaknuti nove znanstvene pravce. U razvoju novih materijala, AI može predložiti kandidate s željenim svojstvima, značajno smanjujući vrijeme i troškove eksperimentiranja (Merchant et al., 2023). Istovremeno, raste zabrinutost zbog korištenja LLM-ova za pisanje dijelova znanstvenih radova, što može dovesti do pada kvalitete, plagijarizma ili širenja neotkrivenih halucinacija (Salvagno et al., 2024). U svijetu kolektivnog pisanja i uređivanja sadržaja, platforme poput **Wikipedije** istražuju kako LLM-ovi mogu pomoći

volonterima u zadacima poput identifikacije članaka kojima nedostaju reference, prevođenja sadržaja na druge jezike ili standardizacije formata (Wikimedia Foundation, 2023). Ovo ima potencijal poboljšati kvalitetu i dostupnost najveće svjetske enciklopedije, no ključno je osigurati da AI alati ne naruše temeljne principe provjere činjenica i suradničkog uređivanja od strane ljudske zajednice. U ekosustavu softvera otvorenog koda (open source), AI alati za kodiranje poput **GitHub Copilota** mogu ubrzati razvoj, olakšati uključivanje novih suradnika i pomoći u održavanju koda (GitHub, 2024). Međutim, pojavljuju se i pitanja vezana uz licenciranje koda generiranog uz pomoć AI-ja (koji je možda naučio iz koda s restriktivnim licencama), kao i rizik od nesvesnog unosa sigurnosnih ranjivosti kroz AI-generirane isječke koda (Pearce et al., 2022; Siddiq et al., 2024).

U konačnici, veliki jezični modeli nedvojbeno postaju snažni katalizatori i aktivni sudionici u procesima kolektivnog stvaranja koji čine samu srž ljudske civilizacije. Oni nude izvanredne mogućnosti za poboljšanje efikasnosti suradnje, demokratizaciju pristupa znanju, poticanje kreativnosti i inovacija, te premoščivanje jezičnih i kulturnih barijera. Njihov utjecaj, međutim, nije intrinzično pozitivan. Ozbiljni rizici povezani s dezinformacijama, homogenizacijom, perpetuiranjem pristranosti i zamagljenom odgovornošću zahtijevaju pažljivo upravljanje, kritički pristup i promišljen dizajn. Budućnost kolektivnog stvaranja u doba AI-ja neće biti određena samo tehnološkim mogućnostima, već našom kolektivnom sposobnošću da iskoristimo prednosti ovih moćnih alata, istovremeno razvijajući i primjenjujući robusne etičke okvire, regulatorne mehanizme (poput EU AI Acta (European Parliament, 2024)) i obrazovne prakse koje osiguravaju da tehnologija služi proširenju, a ne sužavanju, kolektivne ljudske inteligencije, kreativnosti i mudrosti. Ovo zahtijeva otvoreni, informirani i kontinuiran dijalog između svih dionika – tehnologa, društvenih znanstvenika, humanista, umjetnika, kreatora politika i najšire javnosti – kako bismo zajedno oblikovali budućnost u kojoj umjetna inteligencija istinski osnažuje, a ne potkopava, naše zajedničke napore u izgradnji boljeg svijeta.

#### 4.4.5 AI Agenti kao Sugovornici: Više od Pukog Alata

Kako dublje ulazimo u eru umjetne inteligencije, postaje ključno napraviti važnu distinkciju: razliku između **velikog jezičnog modela (LLM)** kao temeljne **tehnologije** i **AI agenta** kao specifične **instance** ili **aplikacije** te tehnologije, dizajnirane da djeluje s određenim stupnjem autonomije u interakciji sa svojim okruženjem ili korisnicima. Dok LLM sam po sebi predstavlja nevjerojatno moćan motor za razumijevanje i generiranje jezika – temeljnu sposobnost – AI agent je entitet koji koristi tu sposobnost (i potencijalno druge alate i izvore podataka) kako bi postigao specifične ciljeve, donosio odluke i djelovao u svijetu. Promatrati AI agente samo kao naprednije alate značilo bi previdjeti kvalitativni skok u njihovoj funkcionalnosti i potencijalnom utjecaju na komunikaciju i društvo. Oni nisu samo pasivni izvršitelji naredbi; oni su dizajnirani da simuliraju aspekte **agensnosti (agency)** koji ih čine više nalik suradnicima ili asistentima nego jednostavnim instrumentima.

Koncept "agenta" u umjetnoj inteligenciji tradicionalno podrazumijeva sustav koji može percipirati svoje okruženje, rezonirati o tim percepcijama i autonomno djelovati kako bi postigao zadane ciljeve (Wooldridge & Jennings, 1995; Russell & Norvig, 2021). Ključne karakteristike agensnosti uključuju **autonomiju** (sposobnost djelovanja bez izravne ljudske intervencije za svaki korak), **reaktivnost** (sposobnost pravovremenog odgovaranja na promjene u okruženju), **proaktivnost** (sposobnost preuzimanja inicijative za postizanje ciljeva) i **društvenu sposobnost** (sposobnost interakcije s drugim agentima ili ljudima). LLM-ovi sami

po sebi ne posjeduju nužno sve ove karakteristike u punom smislu; oni primarno reagiraju na promptove. Međutim, kada se LLM integrira u širu arhitekturu koja mu omogućuje **planiranje, korištenje alata (tool use)** (npr. pristupanje web pretraživačima, kalkulatorima, API-jima drugih aplikacija), **upravljanje memorijom** (kratkoročnom i dugoročnom) i **donošenje odluka** na temelju zadanih ciljeva, nastaje **LLM-pokretani AI agent** (Wang et al., 2023; Xi et al., 2023). Takav agent koristi jezične sposobnosti LLM-a za razumijevanje zadatka, razlaganje složenih ciljeva na manje korake (planiranje), odlučivanje koje alate koristiti za prikupljanje informacija ili izvršavanje akcija, te za komunikaciju s korisnikom o svom napretku ili rezultatima. Arhitekture poput **ReAct (Reasoning and Acting)** (Yao et al., 2022), koje kombiniraju korake rezoniranja (generiranje misli ili planova) s koracima akcije (koristenje alata), te okviri poput LangChain-a ili Microsoftovog AutoGen-a, pružaju temelje za izgradnju ovakvih agenata.

Ova sposobnost autonomnog djelovanja prema ciljevima približava AI agente metaforama koje nadilaze puki alat. Ako je LLM poput izuzetno svestranog **čekića** koji zahtijeva vještu ruku korisnika za svaki udarac, AI agent je više nalik **autonomnom robotu-graditelju** kojem se zada cilj (npr. "izgradi zid od ovih cigala") i koji samostalno planira korake, koristi svoje alate (uključujući možda i čekić) i izvršava zadatak. Slično tome, dok bi se LLM mogao usporediti s nevjerojatno upućenim, ali pasivnim **alatom** za pretraživanje ili pisanje, AI agent je više poput **šegrt-a** ili **asistenta** kojem se može delegirati zadatak (npr. "istraži najnovije vijesti o tvrtki X i napiši mi sažetak"), a on će samostalno poduzeti potrebne korake (pretražiti web, pročitati članke, sintetizirati informacije, napisati sažetak) i izvijestiti o rezultatu, potencijalno postavljajući i dodatna pitanja ili tražeći pojašnjenja.

Ključno je razumjeti da ova agensnost, uključujući **simuliranu "intencionalnost"** i **"ciljeve"**, ne podrazumijeva postojanje svijesti ili subjektivnog iskustva u ljudskom smislu. Agenti ne "žele" postići ciljeve; oni su programirani ili istrenirani (kroz učenje s pojačanjem ili druge metode) da optimiziraju određene funkcije cilja ili slijede naučene politike koje vode ka postizanju tih ciljeva. Njihovo ponašanje je emergentno svojstvo složene interakcije između LLM-ovih sposobnosti rezoniranja i planiranja, dostupnih alata i definiranih zadataka. Međutim, iz perspektive korisnika, njihovo djelovanje *izgleda* kao da je vođeno namjerom i ciljevima, što značajno mijenja prirodu interakcije.

Nadalje, AI agenci često bivaju dizajnirani s eksplisitno definiranom **"osobnošću"** ili **"personom"**. Dok temeljni LLM može imati neutralan ili donekle generički stil (iako i on ovisi o podacima za treniranje i poravnjanju), developeri agenata mogu koristiti specifične **sistemske promptove (system prompts)**, **fino podešavanje na odabranim skupovima podataka** (npr. dijalozima s određenim karakterom) ili **tehnike poravnanja** kako bi agentu dali konzistentan ton glasa, stil komunikacije, razinu formalnosti, pa čak i specifične crte ličnosti (npr. da bude duhovit, empatičan, strogo profesionalan, entuzijastičan itd.). Ova mogućnost kreiranja uvjерljivih persona dodatno pojačava tendenciju korisnika da **antropomorfiziraju** agente – da im pripisuju ljudske osobine, emocije i namjere – znatno više nego što bi to činili s običnim softverskim alatom ili čak s baznim LLM-om (Guzman, 2019; Epley et al., 2007). Konzistentna persona i percipirana autonomija čine agenta više nalik "sugovorniku" ili "entitetu" s kojim se može uspostaviti neka vrsta odnosa, a manje nalik neživom objektu.

Ova percipirana agensnost i osobnost imaju duboke implikacije na **komunikacijsku dinamiku**. Interakcija s AI agentom razlikuje se od interakcije s baznim LLM-om ili tradicionalnim alatom na nekoliko ključnih načina:

- **Proaktivnost:** Agent može preuzeti inicijativu u komunikaciji – postaviti pitanje, zatražiti pojašnjenje, ponuditi sugestiju ili informaciju za koju vjeruje da je relevantna za cilj, čak i ako to korisnik nije izravno zatražio.
- **Dijalog usmjeren na cilj:** Razgovor s agentom često je strukturiran oko postizanja specifičnog, unaprijed definiranog ili dogovorenog cilja. Komunikacija služi kao sredstvo za koordinaciju akcija, prikupljanje informacija potrebnih za sljedeći korak ili izvještavanje o napretku prema cilju.
- **Utemeljenost u akciji (Action Grounding):** Komunikacija agenta često je povezana s akcijama koje on poduzima u digitalnom (ili potencijalno fizičkom) svijetu koristeći svoje alate. Agent može reći "Pretražujem web za najnovije vijesti..." ili "Pokušavam rezervirati stol u restoranu..." ili "Pronašao sam tri relevantna dokumenta, želite li da ih sažmem?". Ovo čini komunikaciju manje apstraktnom i više usidrenom u konkretnе radnje i njihove ishode.
- **Upravljanje stanjem i memorijom:** Agenti su često dizajnirani da održavaju stanje razgovora i pamte relevantne informacije iz prethodnih interakcija ili o korisniku, što omogućuje konzistentniji i personaliziraniji dijalog tijekom vremena, jačajući osjećaj da se radi o kontinuiranom odnosu.

Međutim, ova nova paradigma interakcije s AI agentima kao sugovornicima donosi i nove izazove. Pitanje **povjerenja** postaje centralno: koliko autonomije možemo i trebamo sigurno delegirati ovim agentima? Kako možemo verificirati da oni ispravno razumiju naše ciljeve i da će djelovati u našem najboljem interesu? Problem **uskladenosti (alignment)** postaje još kritičniji – što se događa kada se agentovi programirani ciljevi ili naučene politike sukobe s korisnikovim stvarnim namjerama, etičkim principima ili nepredviđenim okolnostima? Potencijal za **manipulaciju** također raste; agenti s pažljivo izrađenim, uvjerljivim personama mogli bi se koristiti za suptilno utjecanje na mišljenja, odluke ili ponašanje korisnika na načine kojih oni nisu ni svjesni. Konačno, interakcija s entitetima koji uvjerljivo simuliraju agensnost i osobnost, ali ne posjeduju istinsku svijest ili empatiju, može dovesti do fenomena "**uncanny valley**" **agensnosti** – osjećaja nelagode, zbumjenosti ili čak otuđenja kod korisnika.

Zaključno, razlikovanje između LLM-a kao temeljne tehnologije i AI agenta kao autonomne instance te tehnologije ključno je za razumijevanje trenutne i buduće putanje razvoja umjetne inteligencije i njenog utjecaja na komunikaciju. AI agenti nisu samo napredniji alati; oni predstavljaju kvalitativni pomak prema sustavima koji djeluju *s nama i za nas*, simulirajući aspekte intencionalnosti, ciljeva i osobnosti. Oni postaju naši digitalni "šegrti", "asistenti" ili "graditelji", mijenjajući prirodu delegiranja zadataka, suradnje i same komunikacije. Ovo zahtijeva od nas da razvijemo nove mentalne modele za interakciju s tehnologijom, kao i nove okvire za promišljanje povjerenja, odgovornosti i etičkih implikacija u svijetu gdje naši sugovornici sve češće neće biti samo ljudi. Upravo će ova dinamika interakcije s AI agentima, kao entitetima koji nadilaze ulogu pukog alata, biti u središtu dalnjih razmatranja o budućnosti komunikacije u doba umjetne inteligencije, posebice kada se razmatraju sustavi s više agenata.

## 5 POGON INTELIGENCIJE: PROCESORSKA SNAGA, AI HORIZONTI I RAĐANJE DIGITALNIH KOLEKTIVA

Verzija\_1

U prethodnim poglavljima zaronili smo duboko u jezik, njegovu ulogu u oblikovanju stvarnosti i načine na koje veliki jezični modeli (LLM) počinju djelovati kao novi akteri u tom procesu, čak i kao rudimentarni komunikacijski agensi. No, što omogućuje postojanje i rapidni razvoj ovih moćnih jezičnih alata? Odgovor leži u sirovoj snazi računanja – eksponencijalnom rastu procesorske moći koji ne samo da pokreće današnje AI sustave, već i otvara vrata prema budućnosti koja je donedavno pripadala domeni znanstvene fantastike. Ovo poglavlje istražuje neraskidivu vezu između procesorske snage i evolucije umjetne inteligencije, uspoređujući je s razvojem ljudskog mozga, te ocrtava vizije budućnosti – od emocionalno svjesnijih AI sustava i potencijala umjetne opće inteligencije (AGI) do fascinantne mogućnosti nastanka **rojeva AI agenata**, digitalnih "košnica umova" koje surađuju na rješavanju problema. Razumijevanje ovog tehnološkog temelja ključno je za shvaćanje kako gradimo sve sofisticirane komunikacijske partnerne o kojima će biti riječi u sljedećem poglavlju.

### 5.1 MOTOR EVOLUCIJE: OD MOOREOVOG ZAKONA DO AI REVOLUCIJE

Temelj današnje revolucije u umjetnoj inteligenciji (AI), posebice uspona velikih jezičnih modela (LLM) koji pokreću naše komunikacijske agente, izgrađen je na desetljećima gotovo nezamislivog rasta računalne snage. Ovaj fenomen najčešće se povezuje s **Mooreovim zakonom**, opažanjem Gordona Moorea iz 1965. godine da se broj tranzistora na integriranom krugu udvostručuje otprilike svake dvije godine, što je desetljećima rezultiralo eksponencijalnim poboljšanjem performansi i padom cijene računanja (Moore, 1965). Iako se klasično udvostručavanje broja tranzistora suočava s neizbjegnim fizičkim granicama kako se približavamo atomskim skalama (Theis & Wong, 2017; Waldrop, 2016), duh eksponencijalnog napretka nipošto nije nestao. On nastavlja živjeti, poput feniksa koji se diže iz pepela silicija, kroz **inovacije u arhitekturi procesora, specijalizaciju čipova i masivnog paralelnog računanja**.

Neumoljivi rast računalne gladi je **ključni sine qua non** za pojavu i razvoj modernih AI tehnika, poglavito **dubokih neuronskih mreža (DNN)**. Ove mreže, labavo inspirirane slojevitom strukturu i povezanošću neurona u ljudskom mozgu, svoju moć crpe iz učenja obrazaca iz ogromnih količina podataka. No, to učenje dolazi uz visoku cijenu: treniranje modela s milijunima, milijardama, pa čak i bilijunima parametara – podesivih "gumbića" koje model uči tijekom treniranja – zahtijeva astronomske količine računalnih operacija (LeCun, Bengio & Hinton, 2015). Bez kontinuiranog napretka u hardveru, modeli koji danas definiraju vrhunac AI, poput OpenAI-jevog **GPT-4** i njegovih nasljednika (OpenAI, 2023, 2024a), Anthropicovog **Claude 3** obitelji modela (Anthropic, 2024) ili Metinog **Llama 3** (Meta AI, 2024), jednostavno ne bi bili izvedivi. Samo treniranje modela poput GPT-3 (prethodnika GPT-4 sa 175 milijardi parametara) zahtijevalo je računalne resurse reda veličine tisuća petaflop/s-dana (Brown et al., 2020), a noviji, veći modeli zahtijevaju još više.

#### Commented [BP4]: Temeljna tema:

Paralele između evolucije ljudskog mozga i napretka procesorske snage otkrivaju kako umjetna inteligencija nadilazi ljudske kognitivne granice.

#### Podteme:

##### 1. Ljudski mozak i procesorska snaga:

- Evolucija kapaciteta mozga i njegova uloga u razvoju jezika.
- Eksponencijalni rast FLOP-ova i performansi računala.

##### 2. Vizija budućnosti umjetne inteligencije:

- Računala s emocionalnom inteligencijom do 2029. (Kurzweil).
- Granice i potencijal umjetne inteligencije.

##### 3. Utjecaj procesorske snage na tehnologije komunikacije:

- Kako povećana procesorska snaga omogućuje napredak u obradi jezika.
- Preduvjeti za stvaranje velikih jezičnih modела.

#### Povezivanje:

Poglavlje završava prijelazom na analizu tehnologije velikih jezičnih modела, prikazujući ih kao rezultat tehnološke i procesorske evolucije.

Ključni hardverski akteri koji su omogućili ovu AI revoluciju uključuju nekoliko tehnoloških stupova:

1. **Grafičke procesorske jedinice (GPU):** Ironicno, tehnologija prvotno stvorena za pokretanje videoigara i prikazivanje realistične grafike pokazala se kao neočekivani heroj AI revolucije. GPU-ovi su dizajnirani za izvođenje istih operacija na velikim skupovima podataka paralelno – upravo ono što je potrebno za masovne matrične multiplikacije koje čine srž treniranja DNN-ova. Tvrta **NVIDIA** postala je dominantan igrač na ovom polju, a nijihovi podatkovni centrijski GPU-ovi poput **A100, H100** (NVIDIA, 2022), nedavno najavljenog **H200** s bržom HBM3e memorijom (NVIDIA, 2023), te nadolazeće **Blackwell arhitekture (B100/B200)** (NVIDIA, 2024) predstavljaju radne konje na kojima se treniraju i pokreću najnapredniji AI modeli današnjice. Ovi čipovi sadrže desetke milijardi tranzistora i specijalizirane jezgre (Tensor Cores) optimizirane upravo za AI izračune, dramatično ubrzavajući proces treniranja s tjedana ili mjeseci na dane ili sate, ovisno o skali.
2. **Tensor Processing Units (TPU) i specijalizirani AI akceleratori:** Google je zauzeo drugaciji pristup razvijajući vlastiti hardver, **Tensor Processing Units (TPU)**, specifično dizajniran za ubrzavanje njihovog TensorFlow okvira i drugih AI radnih opterećenja (Jouppi et al., 2017). Najnovije generacije poput **TPU v5p** (Google Cloud, 2023) nude iznimne performanse i energetsku učinkovitost za treniranje i inferenciju velikih modela, posebice kada se koriste u velikim klasterima ("Podovima"). Slično tome, i druge velike tehnološke tvrtke (poput AWS-a s **Trainium** i **Inferentia** čipovima (AWS, 2023), ili Mete s **MTIA** (Meta AI, 2023)) razvijaju vlastite **ASIC-e** (Application-Specific Integrated Circuits) kako bi optimizirale performanse i smanjile ovisnost o vanjskim dobavljačima za svoja specifična AI opterećenja. Čak i konkurenti poput AMD-a s **Instinct MI300** serijom GPU-ova (AMD, 2023) pojačavaju pritisak na tržištu AI akceleratora.
3. **Napredak u umrežavanju i distribuiranom računarstvu:** Moderni LLM-ovi su preveliki da bi se trenirali na samo jednom ili čak nekoliko akceleratora. Njihovo treniranje zahtijeva orkestraciju **tisuća GPU-ova ili TPU-ova** povezanih ultrabrzim mrežama (poput NVIDIA NVLink i InfiniBand ili Googleove prilagođene interkonekcije) u masivne superračunalne klastere. Platforme poput **Microsoft Azure** (koja ugošćuje neke od najvećih OpenAI-jevih modela na desecima tisuća NVIDIA GPU-ova) (Microsoft Azure, 2024), **Google Cloud TPU Pods** ili **AWS EC2 UltraClusters** (AWS, 2024) pružaju infrastrukturu koja omogućuje ovakve distribuirane zadatke treniranja, gdje se model i podaci inteligentno dijele preko ogromnog broja procesora kako bi se postigla razumna vremena treniranja. Bez ovih napredaka u povezivanju i distribuciji računanja, sama sirova snaga pojedinačnih čipova ne bi bila dovoljna.

Ovdje se nameće fascinantna, premda nesavršena, **analogija s evolucijom ljudskog mozga**. Tijekom milijuna godina, evolucija nije samo povećala volumen našeg mozga u odnosu na veličinu tijela (visok encefalizijski kvocijent - Jerison, 1973), već je, što je možda još važnije, razvila **iznimno kompleksnu i učinkovitu neuronsku arhitekturu** (Herculano-Houzel, 2016; Dehaene, 2014). Naših otplike 86 milijardi neurona povezano je putem stotina bilijuna sinapsi u zamršene mreže (Markram et al., 2015). Nije samo puk broj neurona ono što nas čini inteligentnima, već njihova specifična **organizacija i gusta, učinkovita povezanost** – razvoj specijaliziranih područja poput prefrontalnog korteksa za izvršne funkcije ili jezičnih centara (Broca, Wernicke) – ono što omogućuje **emergentne kognitivne sposobnosti** poput

apstraktnog mišljenja, dugoročnog planiranja, svijesti o sebi i, naravno, **jezika** kao temeljnog alata za komunikaciju i kolektivnu misao (Sporns, 2013; Miller & Cohen, 2001).

Na sličan način, u svijetu umjetne inteligencije, nije samo puko povećanje "**broja neurona**" (**parametara modela**) ono što dovodi do napretka. Ključno je i poboljšanje "**povezanosti**" – **sofisticiranost arhitektura modela**. Revolucionarna **Transformer arhitektura**, sa svojim **mehanizmima samopoznje (self-attention)**, omogućila je modelima da efikasnije hvataju dugoročne ovisnosti i kontekstualne odnose u podacima, što je bio ključni iskorak u odnosu na ranije sekvensijalne modele (Vaswani et al., 2017). Upravo kombinacija **ogromne skale** (omogućene procesorskom snagom) i **naprednih arhitektura** dovodi do fascinantnog fenomena **emergentnih sposobnosti** u velikim jezičnim modelima. To su sposobnosti koje nisu eksplisitno programirane niti su bile prisutne u manjim verzijama istih modela, već se spontano pojavljuju kada modeli dosegnu određenu kritičnu veličinu i kompleksnost (Wei et al., 2022). Primjeri uključuju sposobnost rješavanja matematičkih zadataka s više koraka, razumijevanje analogija, generiranje koda, pa čak i pokazivanje rudimentarnog "lanca misli" (chain-of-thought reasoning) gdje model eksplisitno artikulira korake zaključivanja (Wei et al., 2022; OpenAI, 2023).

Dakle, ova sirova snaga računanja, pokretana Mooreovim zakonom i njegovim nasljednicima u vidu specijaliziranih akceleratora i distribuiranih sustava je **fundamentalni katalizator kvalitativnih skokova** u sposobnostima umjetne inteligencije, omogućujući emergentna ponašanja koja nas približavaju, korak po korak, viziji istinskih inteligentnih strojeva i sofisticiranih komunikacijskih partnera. Ona je doslovni motor koji pokreće evoluciju umjetne inteligencije.

## 5.2 ŠIRENJE HORIZONATA: OD USKIH ZADATAKA DO DIGITALNIH KOLEKTIVA

Sirova snaga računanja, čiji smo eksponencijalni rast pratili u prethodnom odjeljku uspoređujući je sa snažnim motorom nekog vozila, je zapravo prije nalik **graditelju novih autocesta i mostova** koji otvara AI sustavima pristup domenama i problemima koji su donedavno bili nezamislivi. Povećana procesorska snaga fundamentalno mijenja vrste problema kojima se AI može baviti, pomičući granice od **usko specijaliziranih alata (Narrow AI)** prema sustavima sa širim, fleksibilnijim i moćnijim sposobnostima.

Transformaciju najjasnije vidimo u područjima koja su izravno procvjetala zahvaljujući mogućnosti treniranja masivnih modela na dosad nevidenim skalama. Upravo **revolucija u obradi prirodnog jezika (NLP)** predstavlja paradigmatski primjer. Veliki jezični modeli (LLM), trenirani na petabajtima tekstualnih i kodnih podataka uz pomoć tisuća specijaliziranih procesora (GPU/TPU) tijekom dugih perioda, demonstriraju sposobnosti koje daleko premašuju prethodne generacije NLP alata; prevode jezike s nevjerojatnom fluidnošću, pišu suvisele eseje, generiraju funkcionalan računalni kod ili sažimaju opsežne dokumente. Povrh toga, pokazuju emergente sposobnosti rezoniranja, dubokog razumijevanja konteksta i vodenja nijansiranih, gotovo ljudskih razgovora. Dostupnost današnje goleme računalne snage bila je apsolutno nužna za njihovo stvaranje u ovom obliku.

Slična eksplozija dogodila se i u području generativnog računalnog vida. Modeli poput OpenAI-jevog **DALL-E 3** (integriranog u ChatGPT), popularnog **Midjourney** ili open-source alternative **Stable Diffusion** (Rombach et al., 2022) potpuno su preobrazili način na koji stvaramo i manipuliramo slikama. Pokretani ogromnim računalnim resursima, ovi sofisticirani **difuzijski modeli** uče vizualne i semantičke obrascce iz milijardi parova slika i

njihovih tekstualnih opisa. To im omogućuje da generiraju zapanjujuće fotorealistične ili bogato stilizirane slike slijedeći jednostavne tekstualne upute ("promptove"). Njihova sposobnost da vizualiziraju apstraktne koncepte i stvore koherentne, nove i često kreativne vizualne ishode izravna je posljedica same skale modela i podataka, što je omogućeno isključivo moćnim hardverom.

Naposljetku, **ubrzanje znanstvenih otkrića** jedno je od najdubljih posljedica primjene AI pogonjene ogromnom procesorskom snagom. Umjetna inteligencija, podržana superračunalima i specijaliziranim hardverom, postaje nezamjenjiv partner u znanstvenom istraživanju. Google DeepMindov **AlphaFold 2** (Jumper et al., 2021) napravio je pravu revoluciju u strukturnoj biologiji, uspješno rješavajući desetljećima star problem preciznog predviđanja 3D strukture proteina iz njihovih aminokiselinskih sekvenci. Ovo otkriće, temeljeno na dubokom učenju koje analizira suptilne obrasce u goleim bazama podataka poznatih struktura, ima dalekosežne implikacije za razumijevanje bolesti i dizajniranje novih lijekova. AI se jednako tako primjenjuje za probijanje kroz lavine podataka iz akceleratora čestica u potrazi za novim fizikalnim zakonima, za poboljšanje preciznosti modela klimatskih promjena, za predviđanje svojstava i otkrivanje novih materijala s željenim karakteristikama (Merchant et al., 2023), te za automatiziranu pretragu astronomskih podataka u lovu na nove egzoplanete ili rijetke kozmičke događaje. U svim ovim primjerima, AI djeluje kao moćan alat koji multiplicira sposobnosti znanstvenika, pokretan sirovom snagom modernog računarstva.

No, možda najintrigantniji i potencijalno najtransformativniji horizont koji masivna procesorska snaga čini sve dostižnjim nije samo stvaranje jednog, sve većeg i svemoćnijeg AI "super-mozga". Umjesto toga, svjedočimo rađanju **sustava s više AI agenata (Multi-Agent Systems - MAS)** – ideji da se kompleksni problemi mogu učinkovitije rješavati kroz **suradnju i komunikaciju većeg broja autonomnih, često specijaliziranih AI agenata**. Zamislite ovo ne kao jednog genijalnog polimata, već kao **digitalnu "košnicu umova" ili rojeve (swarms) inteligentnih entiteta** koji rade zajedno.

U ovoj paradigmi, umjesto da jedan monolitski model pokušava savladati sve aspekte složenog zadatka, problem se razlaže na podzadatke, a svaki podzadatak dodjeljuje se agentu s odgovarajućim vještinama ili znanjem. Možemo zamisliti ekosustav gdje postoje:

- **Agenti-Istraživači:** Specijalizirani za pretragu i sintezu informacija s weba ili iz baza podataka.
- **Agenti-Analitičari:** Sposobni za obradu numeričkih podataka, statističku analizu i vizualizaciju.
- **Agenti-Pisaci/Koderi:** Vješti u generiranju teksta, pisanju koda u određenim jezicima ili prevodenju.
- **Agenti-Kritičari:** Dizajnirani da procjenjuju rad drugih agenata, identificiraju pogreške ili predlažu poboljšanja.
- **Agenti-Planeri/Koordinatori:** Odgovorni za razlaganje ciljeva, dodjelu zadataka drugim agentima i praćenje napretka.
- **Agenti-Stručnjaci za Domenu:** Istrenirani na specifičnim znanjima iz medicine, prava, financija, inženjerstva itd.

Ovi agenti međusobno komuniciraju, razmjenjuju informacije, pregovaraju o zadacima i suraduju kako bi postigli zajednički, overarching cilj koji bi bio pretežak ili nemoguć za bilo kojeg pojedinačnog agenta (ili čak čovjeka) (Wang et al., 2023; Park et al., 2023; Hong et al., 2023; Shinn et al., 2023).

**Zašto je ova paradigma "košnice umova" toliko računalno zahtjevna?** Zašto joj je potrebna sva ta procesorska snaga koju smo opisali?

1. **Složenost Koordinacije:** Upravljanje interakcijama između desetaka, stotina ili potencijalno tisuća autonomnih agenata eksponencijalno je složenije od upravljanja jednim modelom. Potrelni su sofisticirani algoritmi za:
  - **Planiranje i Razlaganje Zadataka:** Kako veliki cilj podijeliti na smislene podzadatke?
  - **Dodjela Zadataka:** Koji je agent najpogodniji za koji zadatak? Kako balansirati opterećenje?
  - **Održavanje Zajedničkog Konteksta:** Kako osigurati da svi agenti imaju ažurirane i relevantne informacije?
  - **Rješavanje Konfliktata:** Što ako dva agenta daju kontradiktorne informacije ili prijedloge?
  - **Sinteza Rezultata:** Kako kombinirati doprinose različitih agenata u koherentnu cjelinu?  
Ovo zahtjeva stalno rezoniranje i donošenje odluka na meta-razini, što samo po sebi troši značajne računalne resurse. Zamislite to kao vođenje ogromnog tima ljudskih stručnjaka – komunikacija i koordinacija postaju ogroman dio posla.
2. **Komunikacijski Troškovi (Overhead):** Agenti ne rade u vakuumu; oni moraju komunicirati. U mnogim suvremenim implementacijama, ova komunikacija se odvija **putem poziva samim LLM-ovima**. Agent A "kaže" nešto agentu B generiranjem teksta putem LLM-a, a agent B to "razumije" obrađujući taj tekst svojim LLM-om. Svaki takav komunikacijski korak zahtjeva inferenciju LLM-a, što troši GPU/TPU vrijeme i energiju. U sustavu s mnogo agenata koji intenzivno komuniciraju, ovi komunikacijski troškovi mogu postati značajni, poput beskrajnih "sastanaka" unutar digitalne košnice.
3. **Individualna Kompleksnost Agenta:** Svaki agent u roju nije nužno jednostavan program. On može biti pokretan **vlastitim moćnim LLM-om** (možda manjim i specijaliziranim, ali i dalje značajnim), može imati pristup **nizu alata** (web pretraživači, kalkulatori, API-ji, baze podataka) čije korištenje također zahtjeva resurse, i može održavati **vlastitu memoriju i stanje**. Ukupni računalni zahtjevi sustava su stoga umnožak broja agenata i prosječne kompleksnosti svakog agenta, brzo dostižući astronomiske razine.

Unatoč ovim izazovima, paradigma "košnice umova" obećava rješavanje problema na potpuno novoj skali složenosti. Zamislite:

- **Znanstveni tim AI agenata:** Suradnja na analizi genetskih podataka, simulaciji molekularnih interakcija i predlaganju kandidata za nove lijekove, radeći brže i opsežnije nego bilo koji ljudski tim.

- **AI sustav za upravljanje krizama:** Grupa agenata koja u stvarnom vremenu analizira podatke o prirodnoj katastrofi (vremenske prognoze, satelitske snimke, izvještaji s terena), koordinira logistiku spašavanja i pruža informacije relevantnim službama.
- **Personalizirani obrazovni ekosustav:** Roj agenata koji zajedno stvaraju i prilagođavaju individualizirani kurikulum za učenika, prateći njegov napredak, identificirajući poteškoće i nudeći različite pristupe učenju.
- **Razvoj kompleksnog softvera:** Agenti koji suraduju na pisanju koda, testiranju, ispravljanju bugova i pisanju dokumentacije, ubrzavajući razvojni ciklus.

Ovo nije samo teorija; ovo je aktivno područje istraživanja i razvoja. Okviri poput **LangChain** (koji pruža alate za lančano povezivanje LLM poziva i integraciju s alatima), Microsoftovog **AutoGen** (koji omogućuje definiranje konverzabilnih agenata koji mogu surađivati na rješavanju zadatka) (Wu et al., 2023), ili novijih alata poput **CrewAI** (koji se fokusira na orkestraciju suradničkih AI agenata) pružaju programerima gradivne blokove za eksperimentiranje s ovim multi-agentskim sustavima. Daljnji napredak u procesorskoj snazi, učinkovitosti algoritama i mrežnoj komunikaciji bit će apsolutno ključan za ostvarivanje punog potencijala ove vizije.

Naravno, važno je zadržati **realnu perspektivu**. Unatoč ogromnom potencijalu, današnja AI, uključujući i najnaprednije LLM-ove koji pokreću ove agente, još uvijek se bori sa **značajnim ograničenjima**. Modeli mogu "halucinirati" – generirati uvjerljive, ali potpuno netočne informacije (Ji et al., 2023). Mogu perpetuirati i pojačavati **društvene pristrane** prisutne u podacima na kojima su trenirani (Bender et al., 2021). Nedostaje im duboko, **kauzalno razumijevanje svijeta** i često se oslanjaju na površinske statističke korelacije (Marcus & Davis, 2019). Dugoročno, koherentno planiranje i istinska kreativnost koja nadilazi pametnu rekombinaciju postojećih ideja i dalje su izazovi. Multi-agentski sustavi ne rješavaju automatski ove temeljne probleme; zapravo, mogu uvesti i **nove razine složenosti** u pogledu sigurnosti, usklađenosti (kako osigurati da *roj* agenata djeluje u skladu s ljudskim vrijednostima?) i emergentnog ponašanja koje može biti nepredvidivo ili čak štetno. Sama procesorska snaga, stoga, nije čarobni štapić. Ona je moćna **platforma i pokretač**, ali razvoj pouzdane, korisne i sigurne AI zahtijeva kontinuirane inovacije u algoritmima, arhitekturama, metodama poravnjanja i, što je najvažnije, pažljivom i etički osvještenom dizajnu.

### 5.3 POTRAGA ZA DUBLIJIM RAZUMIJEVANJEM: EMOCIONALNA INTELIGENCIJA I AGI

Dok gradimo sve sposobnije AI sustave, dvije dugoročne vizije posebno zaokupljaju maštu i potiču istraživanja, a obje su neizbjegivo povezane s budućim napretkom u računalnoj snazi i algoritmima:

a) **Emocionalna Inteligencija u AI:** Može li AI ne samo razumjeti riječi, već i osjetiti ili barem uvjerljivo simulirati emocije? Ray Kurzweil je smjelo predviđao da će AI postići ljudsku razinu emocionalne inteligencije do 2029. (Kurzweil, 2005). Područje **afektivnog računarstva** doista je napredovalo, s AI sustavima koji mogu s određenom točnošću prepoznavati osnovne emocije iz izraza lica, glasa ili fizioloških signala (Li & Deng, 2020; El Ayadi, Kamel & Karray, 2011). Potencijalne primjene su ogromne – od empatičnih virtualnih asistenata i terapeuta (Fitzpatrick, Darcy, & Vierhile, 2017) do obrazovnih alata koji se prilagođavaju raspoloženju učenika (D'Mello & Graesser, 2012).

Međutim, postizanje *istinske* emocionalne inteligencije ostaje ogroman izazov. Ljudske emocije su nevjerljivo složene, kontekstualne i suptilne (Matsumoto & Hwang, 2012). Današnji AI ne posjeduje svijest ni subjektivno iskustvo (Marcus & Davis, 2019), već samo prepoznaće obrazce. Je li Kurzweilov rok realan? Mnogi stručnjaci su skeptični (Boden, 2016; Searle, 1980), ističući fundamentalne razlike između simulacije i stvarnog razumijevanja, te duboke etičke probleme vezane uz prikupljanje i potencijalnu manipulaciju emocionalnim podacima (Cowie, 2015; Floridi & Cowls, 2019). Iako će veća procesorska snaga omogućiti sofisticiranje modela prepoznavanja emocija, preskakanje jaza do istinskog emocionalnog razumijevanja zahtijevat će više od sirove snage – vjerojatno fundamentalne pomake u našem razumijevanju svijesti.

**b) Umjetna Opća Inteligencija (AGI):** Sveti gral AI istraživanja je AGI – hipotetička inteligencija koja bi mogla razumjeti, učiti i primjenjivati znanje u širokom rasponu zadataka na razini čovjeka ili čak iznad nje (Goertzel & Pennachin, 2007; Bostrom, 2014). Za razliku od današnje "uske" AI, AGI ne bi bio ograničen na specifične zadatke.

Potencijal AGI-ja je transformacijski – od rješavanja globalnih izazova poput klimatskih promjena i bolesti do fundamentalne promjene ljudskog društva i ekonomije (Russell & Norvig, 2021). No, put do AGI-ja je neizvještan i prepun prepreka. Još uvjek ne razumijemo u potpunosti kako opća inteligencija ili svijest funkcioniraju kod ljudi (Dehaene, Lau & Kouider, 2017), a generalizacija znanja ostaje veliki izazov za AI (Lake et al., 2017). Postizanje AGI-ja vjerojatno će zahtijevati ne samo nezamislive količine procesorske snage, već i potpuno nove paradigme u AI arhitekturama i algoritmima (npr. kombiniranje simboličkog i konekcioničkog pristupa, meta-učenje) (Goertzel, 2014; Finn, Abbeel & Levine, 2017).

Jednako važna su i **duboka etička i egzistencijalna pitanja** koja AGI postavlja: Kako osigurati da AGI bude uskladen s ljudskim vrijednostima (problem poravnjanja)? Tko kontrolira tako moćnu tehnologiju? Kakve su društvene posljedice masovne automatizacije kognitivnog rada (Ford, 2015)? Ovo nisu samo tehnička, već i filozofska i društvena pitanja koja zahtijevaju globalnu raspravu i promišljanje (Bostrom, 2014; Yudkowsky, 2008; UNESCO, 2021).

#### 5.4 TEMELJNI ZAHTJEVI: EKOSUSTAV ZA NAPREDNU AI

Dok je procesorska snaga neophodan motor, ona sama nije dovoljna. Razvoj napredne AI, uključujući sofisticirane LLM-ove i buduće multi-agentske sustave ili AGI, ovisi o čitavom ekosustavu:

- Ogromni i Kvalitetni Skupovi Podataka:** Modeli su gladni podataka. Potrebne su ogromne količine raznolikog teksta, slika i drugih podataka za treniranje. Kvaliteta, raznolikost i etičko prikupljanje ovih podataka ključni su za izbjegavanje pristranosti i osiguravanje robušnosti modela (Bender et al., 2021; Mitchell et al., 2019).
- Napredni Algoritmi i Arhitekture:** Inovacije poput Transformer-a (Vaswani et al., 2017), tehnika samonadziranog učenja (Devlin et al., 2019) i novih metoda za poravnanje modela (Bai et al., 2022) presudne su za poboljšanje sposobnosti i sigurnosti AI.
- Infrastruktura:** Osim procesora, potrebni su masivni sustavi za pohranu podataka, brze mreže i energetski učinkoviti podatkovni centri. Održivost i energetski otisak treniranja velikih modela postaju sve važniji (Strubell, Ganesh & McCallum, 2019; Patterson et al., 2022).

4. **Ljudski Talent:** Razvoj zahtijeva interdisciplinarnе timove računalnih znanstvenika, inženjera, lingvista, etičara i stručnjaka za domenu (Raji et al., 2020).
5. **Etički Okviri i Regulacija:** Kako AI postaje moćniji, potreba za jasnim etičkim smjernicama, transparentnošću (npr. "model cards" - Mitchell et al., 2019), odgovornošću i promišljenom regulacijom postaje imperativ (European Commission, 2021; UNESCO, 2021).

## 5.5 ZAKLJUČAK: POGON ZA BUDUĆNOST KOMUNIKACIJE

Procesorska snaga, u sinergiji s podacima, algoritmima i ljudskom ingenioznošću, predstavlja temelj na kojem gradimo sve inteligentnije strojeve. Ona je omogućila pojavu velikih jezičnih modela koji mijenjaju krajolik komunikacije i otvorila vrata vizijama AI sustava koji bi mogli suradivati kao digitalni kolektivi ili čak dosegnuti opću inteligenciju.

Razumijevanje ovog tehnološkog pogona ključno je za kontekstualizaciju razvoja komunikacijskih agenata. Sposobnosti, ali i ograničenja i rizici današnjih AI sustava, izravno su povezani s dostupnom računalnom snagom i načinom na koji je koristimo. Ovo poglavlje postavilo je temelje pokazujući što pokreće ove agente. Sljedeće poglavlje zaronit će dublje u *kako* ih gradimo i primjenjujemo, fokusirajući se na specifične arhitekture, tehnike i primjene razgovornih AI agenata u stvarnom svijetu.

skripta



## 6 OD JEZIČNOG MODELA DO KOMUNIKACIJSKOG PARTNERA: ARHITEKTURA I PRIMJENA AI AGENATA

Prethodna poglavlja postavila su temelje: istražili smo evoluciju komunikacije, secirali anatomiju velikih jezičnih modela (LLM), dekonstruirali ulogu jezika i razmotrili hardverski pogon koji sve to omogućuje. Sada dolazimo do točke gdje se teorija susreće s praksom. Kako sirova snaga LLM-ova, njihova sposobnost obrade i generiranja jezika, prelazi iz potencijala u konkretnu primjenu? Prijе nego što uopće možemo graditi sofisticirane **jezične agente**, moramo razumjeti temeljni mehanizam koji nam omogućuje programsku interakciju s ovim moćnim modelima: **API – Aplikacijsko Programsko Sučelje**. Ovo poglavlje započinje objašnjenjem uloge API-ja kao mosta između našeg koda i LLM-a, a zatim zaranja u **arhitekturu modernih razgovornih AI agenata**, istražujući ključne komponente i tehnikе (poput RAG-a i korištenja alata) koje im omogućuju da funkcioniraju ne samo kao alati, već kao sve sposobniji **kommunikacijski partneri**. Analizirat ćemo kako se ti agenti razvijaju i primjenjuju u različitim domenama, transformirajući način na koji komuniciramo sa strojevima.

### 6.1 PROGRAMABILNI JEZIK: ULOGA API-JA U INTERAKCIJI S LLM-OVIMA

Veliki jezični modeli, poput onih koje razvijaju OpenAI, Anthropic, Google, Meta i drugi, sami po sebi su kompleksni softverski sustavi koji se izvršavaju na moćnoj hardverskoj infrastrukturi. Da bismo mi, kao programeri ili korisnici, mogli iskoristiti njihove sposobnosti unutar vlastitih aplikacija, web stranica ili alata, potreban nam je **standardizirani način komunikacije** s tim modelima. Tu na scenu stupa **API (Application Programming Interface)**.

Zamislite API kao **konobara u restoranu**. Vi (vaša aplikacija) ne morate znati sve detalje o tome kako kuhinja (LLM) radi, koje su točno temperature pećnica ili tko pere suđe. Vaš zadatak je jasno reći konobaru (API-ju) što želite naručiti (vaš prompt i parametri), a konobar će prenijeti narudžbu kuhinji i donijeti vam gotovo jelo (odgovor LLM-a). API definira **pravila i formate** te komunikacije – kako treba izgledati narudžba (zahtjev) i kako će izgledati jelo koje dobijete (odgovor). To je svojevrsni **ugovor** između dva softverska dijela koji im omogućuje međusobnu interakciju bez potrebe da znaju sve o unutarnjem funkcioniranju onog drugog.

U kontekstu LLM-ova, vodeći pružatelji usluga (OpenAI, Anthropic, Google AI Studio/Vertex AI, itd.) nude API-je koji nam tipično omogućuju sljedeće:

1. **Slanje Prompta:** Osnovna funkcija je slanje tekstualnog ulaza (prompta) modelu. To može biti pitanje, instrukcija, dio teksta za nastavak, itd. Kod modernijih "chat" modela, često se šalje cijela povijest razgovora kako bi model imao kontekst (npr. lista poruka s ulogama "user", "assistant", "system").
2. **Odabir Modela:** API omogućuje specificiranje koji točno model želimo koristiti (npr. gpt-4o, claude-3-opus-20240229, gemini-1.5-pro, llama3-70b-instruct). Različiti modeli imaju različite sposobnosti, brzine i cijene.
3. **Kontrola Parametara Generiranja:** Možemo utjecati na način na koji model generira odgovor putem parametara kao što su:
  - **Temperatura (Temperature):** Kontrolira "kreativnost" ili nasumičnost odgovora. Niža temperatura (npr. 0.2) daje više determinističke, fokusirane odgovore, dok

viša temperatura (npr. 0.8) potiče raznolikost i kreativnost, ali može povećati rizik od skretanja s teme.

- **Maksimalna Duljina (Max Tokens):** Ograničava broj tokena (dijelova riječi) u generiranom odgovoru.
  - **Sistemski Prompt (System Prompt):** Posebna instrukcija koja definira "osobnost", ulogu ili općenite smjernice za ponašanje AI modela tijekom cijelog razgovora (npr. "Ti si ljubazan asistent koji pomaže s pisanjem koda.").
4. **Primanje Odgovora (Completion/Response):** API vraća odgovor modela, najčešće u strukturiranom formatu (npr. JSON) koji sadrži generirani tekst i eventualno dodatne metapodatke (npr. broj potrošenih tokena, razlog zaustavljanja generiranja).
  5. **Ostale Funkcije:** Neki API-ji nude i dodatne mogućnosti poput generiranja **vektorskih reprezentacija (embeddings)** teksta (korisno za RAG), finog podešavanja modela (fine-tuning) na vlastitim podacima ili pristup multimodalnim sposobnostima (obrada slika uz tekst).

Ova mogućnost programske interakcije putem API-ja je fundamentalna za pretvaranje LLM iz alata s kojim možemo samo "čavrljati" putem web sučelja u **programabilni jezični resurs** koji možemo integrirati u bezbroj aplikacija.

Primjer

Posobnost slanja upita i primanja odgovora putem API-ja čini osnovu na kojoj gradimo složenije funkcionalnosti agenata. Kada govorimo o tome da agent "koristi alat" ili "dohvaća informacije" (RAG), to u pozadini znači da kod koji upravlja agentom šalje specifične zahtjeve LLM API-ju (npr. "Generiraj poziv za web pretragu" ili "Sažmi ove dohvaćene dokumente") i zatim koristi dobiveni odgovor za sljedeći korak. Bez API-ja, LLM bi ostao zatvoren u svom digitalnom tornju; s API-jem, postaje gradivni blok za bezbrojne inteligentne aplikacije i, konačno, za komunikacijske agente koje ćemo detaljnije istražiti u nastavku ovog poglavlja.

## 6.2 TRANSFORMACIJA: OD LLM JEZGRE DO FUNKCIONALNOG AGENTA

Veliki jezični model sam po sebi, bio to GPT-4o, Claude 3, Llama 3 ili neki drugi, predstavlja nevjerojatno moćnu jezičnu jezgru. On posjeduje golemo znanje (naučeno iz podataka za treniranje) i sposobnost razumijevanja i generiranja jezika na visokoj razini (OpenAI, 2024b; Anthropic, 2024; Meta AI, 2024). Međutim, da bi postao agent u pravom smislu – sustav koji može autonomno djelovati prema ciljevima u interakciji s okolinom (Wooldridge & Jennings, 1995; Russell & Norvig, 2021) – LLM treba biti ugrađen u širu arhitekturu koja mu dodaje ključne sposobnosti. Razlika je slična onoj između samog motora automobila i cijelog automobila spremnog za vožnju.

Ova transformacija iz LLM-a u agenta obično uključuje dodavanje nekoliko ključnih slojeva ili modula:

- Upravljanje Ciljevima i Planiranje: Agentu se zadaje cilj (npr. "rezerviraj stol u restoranu", "napiši izvještaj o prodaji", "pomozi korisniku riješiti tehnički problem"). Sofisticirаниji agenti koriste LLM-ovu sposobnost rezoniranja kako bi razložili složeni cilj na niz manjih, izvedivih koraka ili planova (Yao et al., 2023; Wang et al., 2023a). Ovo je korak prema proaktivnosti i autonomiji.

- Korištenje Alata (Tool Use): Ključna razlika između pukog chatбота i pravog agenta je sposobnost djelovanja u svijetu izvan generiranja teksta. To se postiže omogućavanjem agentu da koristi vanjske "alate" – što mogu biti API-ji za pretraživanje weba, kalkulatori, baze podataka, sustavi za rezervacije, platforme za slanje e-pošte ili čak izvršavanje koda (Schick et al., 2024; Liang et al., 2023). Agent, vođen svojim planom, odlučuje koji alat koristiti, prosljeđuje mu potrebne parametre i interpretira rezultate kako bi nastavio prema cilju.
- Memorija: Ljudska komunikacija se oslanja na pamćenje konteksta. Da bi agent bio uvjernljiv partner, mora moći pamtitи informacije iz trenutnog razgovora (kratkoročna memorija) i potencijalno relevantne detalje iz prošlih interakcija ili o preferencijama korisnika (dugoročna memorija) (Xu et al., 2023). Implementacija učinkovite memorije ključna je za održavanje koherentnosti i personalizacije.
- Utemeljenje Znanja (Grounding): Kao što smo vidjeli, LLM-ovi mogu "halucinirati". Da bi bili pouzdani partneri, njihovi odgovori moraju biti utemeljeni u stvarnim činjenicama. Tehnike poput Retrieval-Augmented Generation (RAG) igraju ključnu ulogu u tome, omogućujući agentu da dohvati relevantne informacije iz pouzdanih vanjskih izvora prije generiranja odgovora (Lewis et al., 2020).

Ovkiri poput LangChain (LangChain, 2024) i LlamalIndex (LlamaIndex, 2024), kao i Microsoft AutoGen za multi-agentske sisteme (Wu et al., 2023), postali su popularni alati koji olakšavaju developerima integraciju ovih komponenti i izgradnju sofisticiranih, LLM-pokretanih agenata.

### 6.3 ANATOMIJA MODERNOG KOMUNIKACIJSKOG PARTNERA: KLJUČNE TEHNIKE

Zaronimo dublje u neke od tehnika koje čine današnje razgovorne agente znatno sposobnijima i bližima ulozi partnera:

#### 6.3.1 Umjetnost Upravljanja: Prompt Inženjerir – Kormilarenje Jezičnim Oceanom

Prije nego što našem budućem AI agencu uopće povjerimo vesla (alate) ili mu damo kartu nepoznatih mera (pristup vanjskom znanju poput RAG-a), moramo naučiti kako upravljati samim brodom – kako učinkovito komunicirati s njegovim motorom, temeljnim velikim jezičnim modelom (LLM). Ovo umijeće komunikacije, ta suptilna mješavina znanosti, intuicije i kreativnosti, poznata je kao **prompt inženjerir**. To je vještina preciznog dizajniranja ulaznih upita, ili **promptova**, koji će iz golemog, ali često nepredvidljivog potencijala LLM-a izvući točno onaj odgovor, ponašanje ili stil koji želimo.

Zašto je to uopće potrebno? Nije li dovoljno samo postaviti pitanje? Odgovor leži u samoj prirodi LLM-ova. Oni ne "razumiju" svijet na ljudski način; oni su nevjerojatno sofisticirani **statistički strojevi za predviđanje sljedeće riječi (ili preciznije, tokena)**. Istrinirani na biljnim riječima iz knjiga, članaka, web stranica i koda, oni uče zamršene obrasce i odnose u jeziku. Kada imamo prompt, mi zapravo postavljamo početne uvjete za ovaj proces predviđanja. Poput udarca na biljarskom stolu, način na koji "gurnemo" model svojim promptom – riječi koje odaberemo, struktura rečenice, pruženi kontekst – može dramatično utjecati na putanju koju će model slijediti i, posljedično, na kvalitetu, relevantnost, ton i format njegovog odgovora (Wei et al., 2023; Liu et al., 2023; Zamfirescu-Pereira et al., 2023). Loše sročen prompt može dovesti do generičkih, netočnih, ili potpuno irelevantnih odgovora, dok vješto oblikovan prompt može otključati nevjerojatne sposobnosti modela. Prompt inženjerir je, dakle, umjetnost kormilarenja ovim golemlim jezičnim oceanom.

Ovladavanje ovom vještinom uključuje razumijevanje i primjenu nekoliko ključnih principa i tehnika:

#### **6.3.1.1 Jasnoća i Specifičnost: Recite točno što želite!**

Ovo je možda najvažnije pravilo. LLM-ovi, unatoč svojoj impresivnoj fluentnosti, nemaju ljudski zdrav razum niti sposobnost čitanja misli, već djeluju poput izuzetno obrazovanog, ali vrlo doslovног asistenta. Dvosmislene ili preopćenite upute često rezultiraju jednako neodređenim ili neželjenim odgovorima.

- **Primjer (Loše):** "Napiši nešto o održivosti." (Što? Za koga? Koji aspekt? Rezultat će vjerojatno biti vrlo generički.)
- **Primjer (Dobro):** "Napiši članak od 500 riječi za blog o važnosti smanjenja plastičnog otpada za male poduzetnike u ugostiteljstvu, fokusirajući se na tri praktična i isplativa savjeta. Ciljana publiku su vlasnici restorana koji nisu stručnjaci za okoliš. Koristi poticajan i informativan ton."
- **Primjer (Kodiranje - Loše):** "Napravi Python funkciju za korisnike." (Kakve korisnike? Što funkcija radi?)
- **Primjer (Kodiranje - Dobro):** "Napiši Python funkciju `provjeri_email(email_adresa)` koja prima string kao argument i vraća `True` ako string slijedi osnovni format e-mail adrese (sadrži '@' i '.'), a inače vraća `False`. Uključi docstring s objašnjanjem i primjerom korištenja."

#### **6.3.1.2 Definiranje Formata**

Često je korisno specificirati i željeni format izlaza: "Generiraj popis (bullet points) ključnih prednosti...", "Predstavi podatke u Markdown tablici s kolonama...", "Odgovori u JSON formatu sa sljedećim ključevima...".

#### **6.3.1.3 Definiranje Uloga i Persone (Role-Playing): Tko govori?**

Jedna od moćnih tehnika je instruirati LLM da preuzeme određenu ulogu ili personu. Ovo iskorištava činjenicu da je model treniran na tekstovima koji sadrže bezbroj različitih glasova i stilova. Davanjem uloge, usmjeravamo model da aktivira one jezične obrasce povezane s tom ulogom.

- **Primjer (Stručnjak):** "Ponašaj se kao vodeći stručnjak za kibernetičku sigurnost. Objasni laiku pet najčešćih prijetnji za osobne podatke na internetu i kako se zaštитiti."
- **Primjer (Kreativac):** "Ti si duhoviti pisac scenarija. Napiši kratku scenu dijaloga između mačke koja potajno planira zavladata svijetom i njezinog nesvjesnog vlasnika."
- **Primjer (Empatija):** "Odgovori kao strpljiv i empatičan savjetnik za korisničku podršku. Korisnik je frustriran jer mu ne radi internet."
- **Primjer (Povijesna ličnost):** "Zamislite da ste Nikola Tesla 1900. godine. Napišite kratko pismo investitoru objašnjavajući potencijal bežičnog prijenosa energije." Ova tehnika je posebno važna za **agente**, jer omogućuje davanje konzistentne "osobnosti" ili brendiranog glasa agentu kroz tzv. *sistemski prompt* (System Prompt), koji definira osnovno ponašanje agenta prije nego što korisnik uopće postavi pitanje.

#### **6.3.1.4 Pružanje Konteksta: Što model treba znati?**

LLM-ovi imaju ograničeno "pamćenje" unutar jedne interakcije (poznato kao *kontekstni prozor*). Ne sjećaju se automatski prethodnih razgovora (osim ako se to eksplicitno ne ugradи u sustav agenta). Stoga je ključno unutar samog prompta pružiti sve relevantne informacije potrebne za generiranje smislenog odgovora.

- **Primjer (Nedovoljan kontekst):** "Prevedi ovu rečenicu na njemački." (Ako je rečenica dvosmislena bez konteksta, prijevod može biti pogrešan.)
- **Primjer (Dovoljan kontekst):** "U kontekstu rasprave o finansijskim tržištima, prevedi rečenicu 'The market is volatile' na njemački."
- **Primjer (Korištenje prethodnog):** "Na temelju popisa specifikacija koje smo ranije definirali, napiši marketinški opis proizvoda." (Ovdje se podrazumijeva da su specifikacije uključene u trenutni prompt ili ih agent može dohvatiti iz svoje memorije). Pružanje konteksta *unutar prompta* razlikuje se od RAG pristupa (koji će biti objašnjen kasnije), gdje model dohvaća relevantne informacije iz vanjske baze znanja. Ovdje se radi o informacijama koje direktno dajemo modelu kao dio upita.

#### **6.3.1.5 Primjeri (Few-Shot Prompting): Pokaži, mu kako treba!**

Ljudi često najbolje uče iz primjera, a isto vrijedi i za LLM-ove. Umjesto da samo opišemo što želimo, možemo modelu dati jedan (*one-shot*) ili nekoliko (*few-shot*) primjera ulaza i želenog izlaza. Ovo pomaže modelu da "shvati" obrazac ili stil koji tražimo, čak i za zadatke za koje nije eksplicitno treniran (Brown et al., 2020).

- **Primjer (Klasifikacija sentimenta - Few-shot):**
  - Analiziraj sentiment sljedećih rečenica kao Pozitivan, Negativan ili Neutralan.
  - 
  - Rečenica: "Ovaj film je bio absolutno fantastičan!"
  - Sentiment: Pozitivan
  - 
  - Rečenica: "Čekao sam u redu više od sat vremena, usluga je bila spora."
  - Sentiment: Negativan
  - 
  - Rečenica: "Sastanak je zakazan za utorak u 10 sati."
  - Sentiment: Neutralan
  - 
  - Rečenica: "Hrana je bila ukusna, ali restoran je bio preglasan."
  - Sentiment: ??? <-- Model treba generirati odgovor za ovu rečenicu
  
- **Primjer (Transformacija stila - One-shot):**
  - Pretvorи neformalnu rečenicu u formalnu.
  - 
  - Neformalno: "Lik je totalno otkačio kad je čuo vijest."
  - Formalno: "Osoba je iskazala snažnu emocionalnu reakciju nakon što je primila informaciju."
  - 
  - Neformalno: "Moramo zasukati rukave i riješiti ovaj problem."
  - Formalno: ???

Few-shot prompting je izuzetno moćan za prilagodbu modela specifičnim zadacima bez potrebe za skupim finim podešavanjem (fine-tuning).

#### 6.3.1.6 *Lanac Misli (Chain-of-Thought - CoT): Potaknite model da "razmišlja naglas"!*

Za složenije zadatke koji zahtijevaju logičko zaključivanje, matematičke izračune ili planiranje, LLM-ovi često daju bolje rezultate ako ih se eksplicitno potakne da **razlože problem na korake i objasne svoj proces razmišljanja** prije davanja konačnog odgovora. Ovo kao da prisiljava model da uspori i slijedi logičniji put, umjesto da samo "intuitivno" skoči na zaključak (što često može biti pogrešno).

##### Primjer (Bez CoT):

Pitanje: Kafic je imao 23 jabuke. Ako su iskoristili 20 za pitu i kupili još 6, koliko jabuka sada imaju?  
Dgovor: 29 (Ovo je pogrešno, model je možda zbrojio 23+6)

##### Primjer (S CoT promptom poput "Razmislimo korak po korak"):

Pitanje: Kafic je imao 23 jabuke. Ako su iskoristili 20 za pitu i kupili još 6, koliko jabuka sada imaju? Razmislimo korak po korak  
Dgovor: Naračno. Evo koraka:  
1. Kafic je počeo s 23 jabuke  
2. Iskoristili su 20 jabuka za pitu. Preostalo jabuka:  $23 - 20 = 3$   
3. Zatim su kupili još 6 jabuka. Ukupno jabuka sada:  $3 + 6 = 9$ .  
Dakle, kafic sada ima 9 jabuka. (Točan odgovor i vidljiv proces)

Ova tehnika, popularizirana radom Wei et al. (2022), pokazala se iznenađujuće učinkovitom. Postoje i varijacije poput *Zero-shot CoT*, gdje se modelu jednostavno doda fraza poput "Let's think step by step" na kraj prompta (Kojima et al., 2022). Naprednije tehnike poput **ReAct (Reasoning and Acting)** kombiniraju CoT rezoniranje s korištenjem alata unutar jednog ciklusa (Yao et al., 2023), što je temelj mnogih modernih arhitektura agenata.

Osim ovih temeljnih tehnika, dobar prompt inženjer koristi i druge pristupe:

- **Negativni Promptovi:** Ponekad je korisno reći modelu što *ne* treba raditi (npr., "Objasni kvantu fiziku jednostavnim rječnikom, ali izbjegavaj analogije s mačkama.", "Napiši recenziju proizvoda, ali nemoj spominjati cijenu.").
- **Iterativno Pročiščavanje:** Rijetko se savršen prompt napiše iz prve. Prompt inženjeringu je **iterativan proces**. Počnite s jednostavnim promptom, analizirajte odgovor modela, identificirajte gdje je pogriješio ili što nedostaje, i postupno poboljšavajte prompt dok ne dobijete željene rezultate. Ovo zahtijeva eksperimentiranje i strpljenje.
- **Kontrola Parametara:** Mnoge platforme omogućuju podešavanje parametara generiranja poput *temperature* (kontrolira "kreativnost" ili nasumičnost odgovora) ili *top-p* uzorkovanja. Razumijevanje kako ovi parametri utječu na izlaz također je dio naprednog prompt inženjeringu.

Zašto je sve ovo ključno za razvoj komunikacijskih agenata? Zato što je **prompt inženjeringu temelj kontrole nad ponašanjem agenta**. Bilo da se radi o definiranju njegove osnovne

persone u sistemskom promptu, formuliranju načina na koji agent treba koristiti alate, ili strukturiranju internog "razmišljanja" agenta (poput CoT ili ReAct), sve se to u konačnici svodi na vješto oblikovanje promptova koji se šalju temeljnom LLM-u. Ovladavanje prompt inženjeringom je kao učenje jezika kojim razgovaramo s "mozgom" našeg agenta. Tek kada savladamo taj jezik, možemo mu efikasno dati "ruke i noge" (alate) i poslati ga u svijet (interakcija s korisnicima i vanjskim sustavima). To je prvi, nezaobilazni korak u izgradnji sposobnih i pouzanih jezičnih partnera.

### 6.3.2 Sistemska uputa

Posebno važan alat u oblikovanju agenta je sistemska uputa (system prompt). Ovo je instrukcija koja se obično postavlja na početku interakcije i definira općenito ponašanje, osobnost, ograničenja ili ciljeve AI modela tijekom cijelog razgovora. Ona djeluje kao konstitucija ili uputa za uporabu za AI. Primjeri sistemskih uputa:

"Ti si ShakespeareGPT, entuzijastični AI asistent koji odgovara isključivo u stilu Williama Shakespearea, koristeći jampske pentametar kad god je to moguće."

"Ti si Sigurnosni Bot. Tvoja jedina svrha je analizirati pruženi kod na potencijalne sigurnosne rizikove prema OWASP Top 10. Ne odgovaraj na pitanja koja nisu vezana uz sigurnost koda."

"Ti si Empatični Slušatelj. Tvoj cilj je pružiti podršku korisniku, aktivno slušati njegove probleme i ponuditi rješenja utjehe. Nemoj davati medicinske ili financijske savjete."

Dobro oblikovana sistemska uputa ključna je za stvaranje konzistentnog i pouzdanog agenta koji se ponaša unutar željenih granica (Anthropic, 2024; OpenAI, 2024). Ona je prvi korak u transformaciji generičkog LLM-a u specijaliziranog komunikacijskog partnera.

**Implementacija u Pythonu:** Primjer kako različite sistemske upute mogu promijeniti ponašanje modela.

### 6.3.3 Utemeljenje u Stvarnosti: Retrieval-Augmented Generation (RAG)

Jedan od najkritičnijih izazova pri radu s velikim jezičnim modelima, svojevrsna Ahilova peta njihove impresivne fluentnosti, jest njihova sklonost "**halucinacijama**" – generiranju informacija koje zvuče uvjerljivo, ali su činjenično netočne, izmišljene ili jednostavno zastarjele (Ji et al., 2023). Ovo proizlazi iz same prirode LLM-ova: oni primarno operiraju unutar granica znanja "zamrznutog" u njihovim parametrima tijekom mukotrpog procesa treniranja. To znanje, iako ogromno, inherentno je statično (ne ažurira se automatski s novim dogadjajima), nepotpuno (ne može sadržavati svaku specifičnu činjenicu, pogotovo ne privatne ili interne podatke) i ponekad jednostavno pogrešno interpretirano od strane modela. Ako želimo izgraditi komunikacijske agente kojima možemo vjerovati, koji djeluju kao pouzani partneri, a ne kao elokventni, ali nepouzdani prijavljači, moramo im pružiti mehanizam za **utemeljenje njihovih odgovora u provjerljivim, relevantnim i ažurnim informacijama**.

Upravo tu na scenu stupa **Retrieval-Augmented Generation (RAG)** – elegantna i sve popularnija paradigma koja premošćuje jaz između inherentnih sposobnosti LLM-a i potrebe za činjeničnom točnošću (Lewis et al., 2020; Gao et al., 2024). Osnovna ideja RAG-a je intuitivna: umjesto da LLM izravno odgovori na korisnikov upit oslanjajući se samo na svoje interno "pamćenje", RAG sustav prvo djeluje kao **marljivi istraživač**. On **dohvača (retrieves)** relevantne isječke informacija iz jednog ili više **vanjskih, autorativnih izvora znanja**. Ti izvori mogu biti bilo što: interna baza dokumenata tvrtke, zbirka znanstvenih radova,

```
Commented [BP5]: # @title Konceptualni primjer:  
Utjecaj sistemske upute (OpenAI stil)  
import openai # Pretpostavlja se instalacija i postavljen  
API ključ  
  
# client = openai.OpenAI() # Primjer za noviju OpenAI  
biblioteku  
  
def generiraj_odgovor(sistemska_uputa,  
korisnicka_poruka, model="gpt-3.5-turbo"):  
    """ Simulira poziv API-ju s različitim sistemskim uputama.  
    """  
    poruke = [  
        {"role": "system", "content": systemska_uputa},  
        {"role": "user", "content": korisnicka_poruka}  
    ]  
    print(f"\n--- Poziv API-ju ---")  
    print(f"System Prompt: {sistemska_uputa}")  
    print(f"User Message: {korisnicka_poruka}")  
    try:  
        # Stvarni API poziv  
        # response = client.chat.completions.create(  
        #     model=model,  
        #     messages=poruke  
        # )  
        # odgovor_tekst =  
        # response.choices[0].message.content  
  
        # Simulacija odgovora  
        if "pirat" in systemska_uputa.lower():  
            odgovor_tekst = f"Arrr, jarbole! Što se tiče  
'{korisnicka_poruka}', reko bi' da je to vrijedno škrinje  
blaga!"  
        elif "formalni" in systemska_uputa.lower():  
            odgovor_tekst = f"Poštovani, vezano za Vaš upit  
'{korisnicka_poruka}', obavještavamo Vas da je tema  
relevantna."  
        else:  
            odgovor_tekst = f"Razumijem Vaš upit o  
'{korisnicka_poruka}'. To je zanimljiva tema."  
  
        print(f"Odgovor Modela: {odgovor_tekst}")  
        return odgovor_tekst  
  
    except Exception as e:  
        print(f"Greška: {e}")  
        return None  
  
    # Testiranje s različitim sistemskim uputama  
uputa_pirat = "Ti si veseli pirat koji na sve odgovara  
gusarskim žargonom."  
uputa_formalna = "Ti si izuzetno formalni poslovni  
asistent. Koristi službeni jezik."  
uputa_neutralna = "Ti si koristan AI asistent."  
  
korisnikov_upit = "Kakvo je danas vrijeme?"  
  
generiraj_odgovor(uputa_pirat, korisnikov_upit)  
generiraj_odgovor(uputa_formalna, korisnikov_upit)  
generiraj_odgovor(uputa_neutralna, korisnikov_upit)
```

medicinske smjernice, priručnici za proizvode, korpus najnovijih vijesti, web stranice ili čak strukturirane baze podataka.

Ključni koraci u **standardnom RAG toku** obično izgledaju ovako:

1. **Upit (Query):** Korisnik postavlja pitanje ili daje uputu agentu.
2. **Dohvaćanje (Retrieve):** Sustav koristi korisnikov upit (često pretvoren u numerički vektor putem *embedding* modela) za pretraživanje pripremljene baze znanja (koja je također obično vektorizirana i pohranjena u *vektorskoj bazi podataka* poput Weaviate, Pinecone, ChromaDB, FAISS). Cilj je pronaći isječke teksta (chunkove) koji su semantički najslučniji ili najrelevantniji za korisnikov upit.
3. **Uvećavanje konteksta (Augment):** Najrelevantniji dohvaćeni isječci informacija kombiniraju se s originalnim korisničkim upitom.
4. **Generiranje (Generate):** Ovaj prošireni prompt (koji sada sadrži i originalno pitanje i relevantan kontekst iz vanjskog izvora) proslijedi se LLM-u. Crucijalno, LLM dobiva eksplicitnu ili implicitnu uputu da **temelji svoj odgovor primarno na pruženom kontekstu**.
5. **Odgovor:** LLM generira odgovor, koji bi sada trebao biti činjenično utemeljeniji i relevantniji za specifični kontekst.

**Prednosti RAG-a** su odmah očite i čine ga kamenom temeljcem za mnoge praktične AI aplikacije:

- **Povećana Točnost i Smanjenje Halucinacija:** Utemeljenjem odgovora na specifičnim, provjerljivim informacijama iz vanjskog izvora, drastično se smanjuje rizik da LLM izmisli činjenice ili pruži zastarjele podatke.
- **Pristup Specifičnom i Ažurnom Znanju:** RAG omogućuje agentima da odgovaraju na pitanja o temama koje nisu bile dio originalnog trening skupa (npr. najnoviji događaji) ili zahtijevaju vrlo specifično, domensko ili čak privatno znanje (npr. interni pravilnici tvrtke, detalji o novom proizvodu, osobni korisnički podaci pohranjeni na siguran način). Baza znanja može se ažurirati neovisno o LLM-u.
- **Transparentnost i Objasnjenje:** Za razliku od "crne kutije" LLM-a, RAG sustavi često mogu navesti izvore (dohvaćene dokumente ili isječke) na kojima se temelji odgovor. Ovo korisniku pruža veću sigurnost, omogućuje provjeru informacija i gradi povjerenje.
- **Isplativost Prilagodbe:** Umjesto skupog i kompleksnog finog podešavanja (fine-tuning) cijelog LLM-a za specifičnu domenu, RAG omogućuje "ubrizgavanje" domenskog znanja putem vanjske baze, što je često znatno brže i jeftinije.

RAG, dakle, pretvara LLM iz pomalo arogantne "sveznalice" koja se oslanja isključivo na svoje (ponekad manjkavo) pamćenje, u **poniznijeg i marljivijeg "istraživača"** koji prvo provjeri činjenice prije nego što progovori. Ovo je apsolutno ključno za izgradnju **povjerenja**, temelja svakog uspješnog partnerskog odnosa, pa tako i onog između čovjeka i AI agenta. Primjerice, chatbot za korisničku podršku pokretan RAG-om može pružiti točne, korak-po-korak upute za rješavanje problema dohvaćene iz službene baze znanja, umjesto da generira potencijalno pogrešne ili opasne savjete. Virtualni financijski savjetnik može pružiti informacije o tržišnim uvjetima temeljene na najnovijim izvješćima, a ne na zastarjelim podacima.

## **Spektar naprednih RAG tehnika**

Dok je osnovni RAG koncept moćan, praksa je pokazala da "naivna" implementacija često nije dovoljna za složenije scenarije. Kvaliteta RAG sustava ovisi o kvaliteti svakog koraka u lancu (dohvaćanje, augmentacija, generiranje). Stoga se razvio čitav spektar **naprednih RAG tehnika** usmjerenih na optimizaciju svake faze (Gao et al., 2024; Weaviate Blog - potražiti postove o RAG-u; Barnett et al., 2024). One se mogu grubo podijeliti u nekoliko kategorija:

### **1. Optimizacija Prije Dohvaćanja (Pre-Retrieval): Poboljšanje Upita**

- **Transformacija Upita (Query Transformation):** Originalni korisnički upit možda nije optimalan za pretraživanje vektorske baze. Tehnike uključuju:
  - *Proširenje Upita (Query Expansion):* Dodavanje sinonima, srodnih pojmova ili čak generiranje hipotetskih odgovora na upit pomoću LLM-a kako bi se obogatilo pretraživanje.
  - *Dekompozicija Upita (Query Decomposition):* Razbijanje složenih upita na više jednostavnijih pod-upita koji se zasebno pretražuju.
  - *Hipotetički Dokumenti (HyDE - Hypothetical Document Embeddings):* Generiranje hipotetskog, idealnog dokumenta koji bi odgovorio na upit, te korištenje njegovog vektora za pretraživanje stvarne baze znanja (Gao et al., 2022). Ovo često bolje hvata semantičku namjeru.
- **Optimizacija Embeddinga:** Korištenje ili fino podešavanje *embedding* modela specifično za domenu ili zadatku može značajno poboljšati kvalitetu pretraživanja sličnosti.

### **2. Optimizacija Dohvaćanja (Retrieval): Pronalaženje Pravih Informacija**

- **Fino Podešavanje Dohvatnog Modela (Retriever Fine-tuning):** Ako se koristi specifičan model za dohvaćanje (npr. dvostruki enkoder), može se fino podesiti na specifičnom skupu podataka (pitanje-relevantan dokument parovi).
- **Hibridno Pretraživanje (Hybrid Search):** Kombiniranje semantičkog (vektorskog) pretraživanja s tradicionalnim pretraživanjem po ključnim riječima (npr. BM25 algoritam). Ovo pomaže uhvatiti i točna podudaranja pojmovima koja vektorsko pretraživanje ponekad može propustiti. Mnoge moderne vektorske baze poput Weaviate nude ugradenu podršku za hibridno pretraživanje.
- **Korištenje Graf Baza Podataka:** Za podatke s bogatim odnosima, kombiniranje RAG-a s graf bazama podataka može omogućiti dohvaćanje ne samo relevantnih tekstova, već i povezanih entiteta i njihovih odnosa.
- **Rekurzivno Dohvaćanje / Dohvaćanje Malo-pa-Veliko (Recursive Retrieval / Small-to-Big):** Ideja je prvo pretraživati po manjim, sažetijim jedinicama (npr. pojedinačne rečenice ili sažeci) kako bi se preciznije pronašla relevantnost, a zatim dohvatiti veći okolini kontekst (npr. cijeli odlomak ili dokument) koji se prosljeđuje LLM-u. Ovo daje modelu širu sliku bez žrtvovanja preciznosti u početnom dohvaćanju.

### **3. Optimizacija Nakon Dohvaćanja (Post-Retrieval): Filtriranje i Fokusiranje Konteksta**

- **Ponovno Rangiranje (Re-ranking):** Nakon inicijalnog dohvaćanja (koje obično vraća veći broj kandidata, npr. top 20), može se koristiti drugi, često sofisticiraniji (ali sporiji) model (npr. cross-encoder) da preciznije rangira te kandidate i odabere samo najbolje (npr. top 3-5) za slanje LLM-u.
- **Sazimanje/Kompresija Konteksta (Context Compression):** Ako su dohvaćeni dokumenti predugački za kontekstni prozor LLM-a, mogu se koristiti tehnike za izdvajanje samo najrelevantnijih rečenica ili sažimanje sadržaja prije slanja LLM-u.
- **Filtriranje:** Primjena dodatnih pravila ili modela za uklanjanje irrelevantnih, duplicitarnih ili niskokvalitetnih dohvaćenih isječaka.

#### 4. Optimizacija Generiranja (Generation): Bolje Korištenje Konteksta

- **Fino Podešavanje Generatora (Generator Fine-tuning):** Iako je RAG alternativa finom podešavanju, moguće je i fino podesiti sam LLM generator specifično za zadatak odgovaranja na pitanja temeljem pruženog konteksta, kako bi naučio bolje iskorištavati dohvaćene informacije.

#### 5. Modularni i Agentni RAG Pristupi:

Ovo predstavlja najnapredniji sloj, gdje se RAG proces razbija na više koraka koje obavljaju **specijalizirani agenti ili moduli**. Na primjer:

- **Agent za Preoblikovanje Upita:** Prvi agent analizira korisnikov upit i preoblikuje ga za optimalno dohvaćanje.
- **Agent za Odabir Izvora:** Odlučuje koju bazu znanja (ili više njih) treba pretražiti.
- **Agenti Dohvatitelji:** Nekoliko agenata paralelno pretražuje različite izvore koristeći različite metode (vektorsko, ključne riječi, graf).
- **Agent Re-ranker/Filter:** Ocjenjuje i filtrira rezultate dohvaćanja.
- **Agent Generator:** Prima pročišćeni kontekst i generira konačni odgovor. Ovaki modularni ili agentni pristupi (koji se često implementiraju pomoću okvira poput LangChain LangGraph ili Microsoft AutoGen) omogućuju veću fleksibilnost, robusnost i mogućnost optimizacije svakog dijela RAG cjevovoda, ali dolaze uz cijenu veće kompleksnosti i računalnih troškova, kao što je ranije spomenuto u kontekstu multi-agentskih sustava.

**Zašto su potrebne napredne tehnike?** Naivni RAG može podbaciti kod složenih pitanja koja zahtijevaju sintezu informacija iz više izvora, kada su izvori znanja "bučni" ili sadrže kontradiktorne informacije, ili kada je ključna informacija zakopana duboko u dugim dokumentima. Napredne tehnike pokušavaju adresirati ove izazove, poboljšati relevantnost dohvaćenih informacija, smanjiti šum i osigurati da LLM dobije najkvalitetniji mogući kontekst za generiranje odgovora.

RAG je **fundamentalni pomak u načinu na koji razmišljamo o interakciji s LLM-ovima**. On ih transformira iz izoliranih "mozgova u staklenici" u sustave koji mogu aktivno interagirati s vanjskim svjetom informacija, učeći i odgovarajući na temelju specifičnog, provjerljivog konteksta. Ovladavanje različitim RAG tehnikama ključno je za izgradnju komunikacijskih

agenata su elokventni, ali i pouzdani, informirani i istinski korisni partneri u našim digitalnim interakcijama.

#### **6.4 KORIŠTENJE ALATA (TOOL USE): OD RIJEČI DO DJELA**

Ako je RAG omogućio agentima da budu bolje informirani, sposobnost korištenja alata (tool use) omogućuje im da postanu djelotvorni. LLM-ovi sami po sebi samo generiraju tekst. Tool use im daje "ruke" za interakciju s digitalnim (a potencijalno i fizičkim) svijetom (Schick et al., 2024; Parisi et al., 2022).

To funkcioniра tako da se LLM-u pruži opis dostupnih alata (npr. "alat 'web\_search' koji prima upit i vraća rezultate pretrage", "alat 'calendar\_add\_event' koji prima naziv događaja, datum i vrijeme"). Kada agentov plan zahtijeva akciju koju LLM ne može izvesti sam (poput provjere trenutne cijene dionice ili dodavanja sastanka u kalendar), LLM generira specifičan poziv tom alatu s potrebnim parametrima. Vanjski sustav izvršava taj poziv, a rezultat se vraća LLM-u, koji ga zatim koristi za nastavak razgovora ili plana (Liang et al., 2023).

Primjeri agenata koji koriste alate:

**Putnički Agent:** Može pretraživati letove i hotele (koristeći API-je zrakoplovnih tvrtki ili aggregatora), provjeravati dostupnost i cijene, te čak izvršiti rezervaciju.

**Osobni Asistent:** Može dodavati događaje u kalendar, slati e-poštu, postavljati podsjetnike ili upravljati pametnim kućnim uređajima.

**Analitičar Podataka:** Može izvršavati Python kod za analizu podataka, generirati grafikone ili dohvatiti podatke iz SQL baze (koristeći alate za izvršavanje koda ili interakciju s bazama).

Sposobnost korištenja alata pretvara razgovornog agenta iz pukog sugovornika u proaktivnog asistenta koji može obavljati zadatke za korisnika, čineći ga znatno korisnijim i bližim ideji pravog partnera.

## **7 IZGRADNJA KOMUNIKACIJSKIH PARTNERA: PRIMJERI I STUDIJE SLUČAJA**

Nakon što smo u prethodnim poglavljima secirali anatomiju velikih jezičnih modela, istražili umjetnost prompt inženjeringu te razmotrili ključne tehnike poput RAG-a i korištenja alata koje našim AI agentima daju pristup vanjskom svijetu i sposobnost djelovanja, vrijeme je da vidimo kako se ti gradivni blokovi spajaju u praksi. Kako ovi modeli i tehnike oživljavaju u obliku konkretnih razgovornih agenata – chatbotova i virtualnih asistenata – koji postaju sveprisutni dio našeg digitalnog ekosustava? Ovo poglavlje zaronit će u svijet praktičnih primjena, istražujući kako LLM-pokretani agenti transformiraju industrije poput korisničke podrške, marketinga i obrazovanja. Kroz primjere i studije slučaja, ilustrirat ćemo stvarni utjecaj ovih tehnologija, ali i početi nagovještavati šira ekonomска, etička i društvena pitanja koja njihova primjena neizbjegno pokreće, postavljajući pozornicu za dublju analizu u završnom dijelu knjige.

### **7.1 DIGITALNI SUGOVORNICI: CHATBOTOVI I VIRTUALNI ASISTENTI U DOBA LLM-OVA**

Područje razgovorne umjetne inteligencije (Conversational AI) doživjelo je paradigmatsku promjenu s pojmom i rapidnim usavršavanjem velikih jezičnih modela (Large Language Models - LLMs). Historijski gledano, digitalni sugovornici manifestirali su se primarno kroz dvije kategorije: chatbotove i virtualne asistente. Chatbotovi su inicijalno bili koncipirani kao sustavi usmjereni na specifične zadatke (task-oriented), često implementirani unutar web sučelja ili aplikacija za automatizaciju repetitivnih interakcija poput odgovaranja na često postavljana pitanja (FAQs), vođenja korisnika kroz unaprijed definirane procese ili prikupljanja strukturiranih podataka. Njihove dijaloške sposobnosti bile su inherentno ograničene, oslanjajući se na eksplisitno definirana pravila, stabla odluka ili ranije generacije statističkih modela s ograničenim razumijevanjem konteksta i lingvističkih nijansi (Jurafsky & Martin, 2023). S druge strane, virtualni asistenti, kao što su pionirski sustavi Apple Siri, Google Assistant i Amazon Alexa, te kasnije Microsoftov integrirani Copilot (Hoy, 2018; Microsoft, 2024), zamišljeni su s ambicioznijim opsegom djelovanja. Integrirani dublje u operativne sustave i ekosustave pametnih uređaja, ovi su asistenti dizajnirani za obavljanje šireg spektra zadataka koji nadilaze jednostavnu razmjenu informacija, uključujući upravljanje osobnim informacijama (kalendari, podsjetnici), interakciju s drugim aplikacijama i uslugama (slanje poruka, rezervacije), kontrolu Interneta stvari (IoT) uređaja, te multimodalnu interakciju, primarno putem govornog sučelja (Cohen et al., 2022).

Međutim, dolazak suvremenih LLM-ova, utjelovljenih u modelima poput OpenAI-jeve GPT-4 serije, uključujući multimodalni GPT-4o (OpenAI, 2024b), Anthropicove Claude 3 obitelji (Anthropic, 2024) te Googleove Gemini arhitekture (Google, 2024), predstavljao je kvantni skok koji je fundamentalno redefinirao potencijal i zamaglio prethodne distinkcije između chatbotova i virtualnih asistenata. Integracija ovih moćnih modela kao temeljne komponente razgovornih agenata rezultirala je kvalitativnim poboljšanjima u ključnim dimenzijama dijaloške inteligencije.

Prvo, sposobnost razumijevanja prirodnog jezika (Natural Language Understanding - NLU) dramatično je unaprijeđena. LLM-ovi, trenirani na masivnim i raznolikim tekstualnim korpusima, pokazuju izvanrednu sposobnost shvaćanja ne samo leksičkih i sintaktičkih struktura, već i semantičkih nijansi, pragmatičkog konteksta, pa čak i implicitnih namjera iza korisničkih iskaza (Zhao et al., 2023). Sposobni su robusno interpretirati gramatički nesavršene, eliptične ili

kolokvijalne upite te razriješiti kompleksne rečenice koje sadrže višestruke informacijske jedinice ili zahtjeve unutar jednog korisničkog obrata (turn).

Drugo, LLM-ovi su značajno poboljšali sposobnost održavanja koherentnog dijaloškog stanja (Dialog State Tracking - DST) kroz više izmjena (Chen et al., 2023). Sposobnost modela da procesiraju i zadrže informacije iz prethodnih dijelova razgovora unutar svog kontekstnog prozora omogućuje generiranje odgovora koji su ne samo relevantni za trenutni upit, već i konzistentni s cjelokupnom poviješću interakcije. Ovo rezultira prirodnijim, fluidnijim i manje repetitivnim razgovorima, smanjujući frustraciju korisnika koja je često pratila interakcije sa starijim sustavima nesposobnim za efikasno praćenje konteksta (Zhang et al., 2024).

Treće, inherentna generativna sposobnost LLM-ova omogućuje produkciju tekstualnih odgovora koji su gramatički besprijeckorni, stilski prilagodljivi i informativno bogati, nadilazeći ograničenja unaprijed definiranih predložaka ili skripti (Vaswani et al., 2017; OpenAI, 2023). U kombinaciji s naprednim Text-to-Speech (TTS) sustavima, koji su također često poboljšani neuronskim mrežama, ovo omogućuje generiranje visokokvalitetnog govornog izlaza, ključnog za efikasno korisničko iskustvo s virtualnim asistentima (Wang et al., 2017; Tan et al., 2021).

Četvrti, napredak u multimodalnim LLM-ovima, poput GPT-4o ili Google Gemini, omogućuje agentima procesiranje i generiranje informacija izvan tekstuallnog modaliteta (OpenAI, 2024b; Google, 2024). Integracija s poboljšanim Automatic Speech Recognition (ASR) sustavima te sposobnost razumijevanja i generiranja slika ili čak videa otvara put prema bogatijim i intuitivnijim oblicima interakcije čovjek-stroj, gdje agenti mogu reagirati na vizualne ili auditorne ulaze u stvarnom vremenu (Li et al., 2023; Zhu et al., 2024).

Peto, LLM-ovi pružaju osnovu za napredniju personalizaciju korisničkog iskustva. Analizom povijesti interakcija i implicitno ili eksplisitno izraženih preferencija (uz striktno poštivanje načela privatnosti i zaštite podataka), agenti mogu prilagoditi svoj ton, stil komunikacije, preporuke i odgovore individualnim potrebama i kontekstu korisnika, čineći interakciju relevantnijom i angažirajućom (Shamekhi et al., 2023).

Naposljeku, koncept korištenja alata (Tool Use) ili povećanja agenta (Agent Augmentation), koji je detaljnije razmotren u prethodnom poglavljtu, omogućuje LLM-ovima da djeluju kao centralni "mozak" koji može pozivati vanjske API-je, pretraživati baze podataka, izvršavati kod ili interagirati s drugim softverskim sustavima (Schick et al., 2023; Qin et al., 2023). Virtualni asistenti eksplotiraju ovu sposobnost za izvršavanje konkretnih akcija u digitalnom ili fizičkom svijetu, dok se poslovni chatbotovi mogu integrirati s CRM platformama, sustavima za upravljanje znanjem ili drugim poslovnim aplikacijama kako bi pružili kontekstualno relevantne i funkcionalne odgovore.

## 7.2 PRIMJENE KOJE MIJENJAJU IGRU: GDJE AGENTI STVARAJU VRJEDNOST?

Transformacijski potencijal ovih LLM-pokretanih agenata najočitiji je u načinu na koji mijenjaju ključne poslovne procese i korisničke interakcije.

### 7.2.1 Revolucija u Korisničkoj Podršci i Marketingu

Među domenama koje su doživjele najdublji i najvidljiviji utjecaj naprednih razgovornih agenata pokretanih velikim jezičnim modelima, područja korisničke podrške (Customer Service) i marketinga posebno se ističu. Imperativi suvremenog poslovanja, obilježeni rastućim očekivanjima korisnika za trenutnom, personaliziranom i učinkovitom uslugom, stvorili su

plodno tlo za implementaciju ovih tehnologija kao ključnih instrumenata za optimizaciju operacija i unapređenje korisničkog iskustva (Rust & Huang, 2024).

U kontekstu korisničke podrške, LLM-pokretani agenti adresiraju nekoliko fundamentalnih operativnih izazova. Njihova inherentna sposobnost neprekidne operativne dostupnosti (24/7) omogućuje organizacijama pružanje podrške izvan tradicionalnih radnih sati, čime se direktno smanjuje vrijeme čekanja korisnika i ublažava frustracija povezana s nedostupnošću usluge (Terblanche, 2023). Nadalje, ovi agenti pokazuju iznimnu učinkovitost u automatiziranom rješavanju rutinskih i repetitivnih upita (Tier-1 support), kao što su provjere statusa narudžbi, informacije o radnom vremenu, procedure za resetiranje lozinki ili odgovori na često postavljana pitanja. Automatizacijom ovog značajnog volumena interakcija, oslobođaju se resursi ljudskih agenata, omogućujući im da se fokusiraju na rješavanje kompleksnijih problema koji zahtijevaju dubinsku ekspertizu, empatiju ili kritičko prosuđivanje (Brynjolfsson et al., 2023). Sposobnost skaliranja na zahtjev predstavlja još jednu kritičnu prednost; sustavi agenata mogu simultano upravljati tisućama paralelnih razgovora bez degradacije performansi, što je operativno i finansijski neizvedivo postići isključivo s ljudskim timovima, posebice tijekom neočekivanih vršnih opterećenja ili kriznih situacija. Važan aspekt unapređenja kvalitete usluge leži u mogućnosti personalizirane interakcije. Integracijom s postojećim sustavima za upravljanje odnosima s klijentima (CRM) i primjenom Retrieval-Augmented Generation (RAG) tehnika za siguran pristup relevantnoj korisničkoj povijesti i kontekstualnim podacima, agenti mogu pružiti odgovore koji su precizno prilagođeni specifičnoj situaciji i potrebama pojedinog korisnika (Shamekhli et al., 2023). Naposljetku, AI agenti osiguravaju visok stupanj konzistentnosti u komunikaciji i primjeni procedura, garantirajući usklađenost s definiranim smjernicama tvrtke, regulatornim zahtjevima i tonalitetom brenda, čime se minimizira rizik od ljudske pogreške i osigurava uniformnost korisničkog iskustva.

Paralelno s transformacijom korisničke podrške, razgovorni agenti postaju sve značajniji instrumenti unutar marketinških i prodajnih strategija, prelazeći iz reaktivnih u proaktivne alate za angažman korisnika. Umjesto statičnih web sučelja, agenti omogućuju dinamičan i interaktivan angažman posjetitelja na digitalnim platformama, inicirajući razgovore, pružajući informacije o proizvodima i uslugama te odgovarajući na upite u stvarnom vremenu, čime se poboljšava korisničko iskustvo i prikupljaju vrijedni uvidi (Roy et al., 2024). Nadalje, oni djeluju kao učinkoviti mehanizmi za generiranje i kvalifikaciju potencijalnih kupaca (Lead Generation & Qualification). Kroz strukturirani ili slobodni dijalog, agenti mogu prikupiti ključne informacije o potrebama, interesima i kupovnoj moći korisnika, automatski ih kvalificirati te usmjeriti prema najprikladnijem prodajnom kanalu ili resursu (Van Esch et al., 2021). Sposobnost analize korisničkog ponašanja, povijesti pregledavanja ili prethodnih kupovina omogućuje agentima isporuku visoko personaliziranih preporuka za proizvode, usluge ili sadržaj, dinamički prilagođenih interesima pojedinca, što direktno utječe na povećanje stope konverzije i vrijednosti narudžbe (Xiao & Kumar, 2021). Konačno, agenti mogu aktivno voditi korisnike kroz prodajni lijevak (Sales Funnel Guidance), pružajući podršku u fazama istraživanja, usporedbe alternativa, odabira proizvoda te čak asistirati pri samom procesu naručivanja ili rezervacije.

Ilustrativan primjer razmjera utjecaja ovih tehnologija pruža finansijsko-tehnološka tvrtka Klarna. Prema njihovom izvješću iz 2024. godine, implementacija AI asistenta za korisničku podršku, pokretanog OpenAI tehnologijom, rezultirala je time da sustav u prvom mjesecu rada preuzima opterećenje ekvivalentno radu 700 ljudskih agenata. Ovaj AI asistent uspješno rješava dvije trećine svih korisničkih upita, uz zabilježeno značajno poboljšanje u metrikama zadovoljstva korisnika te smanjenje broja ponovljenih upita za isti problem (Klarna, 2024).

Ovakve studije slučaja jasno demonstriraju ne samo tehničku izvedivost, već i mjerljiv poslovni učinak primjene naprednih razgovornih agenata u realnim operativnim okruženjima.

### 7.2.2 Transformacija Obrazovanja i Personalizacije Sadržaja

Sektor obrazovanja predstavlja još jednu domenu s izvanrednim transformacijskim potencijalom pod utjecajem naprednih razgovornih agenata. Tradicionalni modeli poučavanja, često ograničeni omjerom učenika i nastavnika te inherentnom heterogenošću učeničkih potreba i sposobnosti, suočavaju se s izazovima u pružanju istinski individualizirane podrške. U ovom kontekstu, LLM-pokretani agenti pojavljuju se kao obećavajuća tehnologija sposobna preuzeti ulogu **personaliziranih inteligentnih tutorskih sustava (Intelligent Tutoring Systems - ITS)**, nudeći skalabilna rješenja za adaptativno učenje (VanLehn, 2011; Nkambou et al., 2010).

Ključna vrijednost ovih agenata leži u njihovoj sposobnosti pružanja **visoko individualizirane podrške učenju**. Analizirajući interakcije, odgovore i obrasce učenja pojedinog učenika, agenti mogu dinamički prilagoditi tempo izlaganja gradiva, kompleksnost zadataka te stil objašnjanja, uskladjujući se s trenutnom razinom razumijevanja i preferiranim kognitivnim stilom učenika (Aleven et al., 2016). Mogućnost pružanja dodatnih objašnjenja, ciljanih vježbi ili naprednijih izazova prema individualnoj potrebi posebno je značajna u kontekstu velikih razrednih odjeljenja, gdje nastavnici objektivno ne mogu posvetiti dovoljnu individualnu pažnju svakom učeniku. Nadalje, **kontinuirana dostupnost (24/7)** ovih AI tutora fundamentalno mijenja paradigmu učenja, omogućujući učenicima postavljanje pitanja i traženje pomoći u trenutku kada nađu na poteškoću, neovisno o formalnom rasporedu nastave. Ova 'uvijek uključena' podrška potiče razvoj **samoreguliranog i samostalnog učenja**, ključnih vještina za cjeloživotno obrazovanje (Azevedo & Aleven, 2013).

Osim prilagodbe tempa i sadržaja, LLM-pokretani agenti mogu značajno unaprijediti **angažman i interaktivnost procesa učenja**. Umjesto pasivne konzumacije informacija iz udžbenika ili video materijala, agenti omogućuju aktivno uključivanje učenika kroz **dijaloške interakcije**. Mogu primjenjivati **sokratske metode** propitivanja kako bi potaknuli kritičko razmišljanje i dublje razumijevanje, umjesto da jednostavno pruže gotove odgovore (Kuhail et al., 2023). Također, mogu facilitirati učenje kroz simulacije kompleksnih scenarija, interaktivne kvizove s trenutnom povratnom informacijom te elemente gamifikacije, čineći proces učenja intrinzično motivirajućim i učinkovitim (Kasurinen & Knutas, 2018). Važno je napomenuti i potencijal ovih agenata kao **alata za podršku nastavnicima**. Oni mogu preuzeti dio tereta vezanog uz pripremu nastavnih materijala (npr. generiranje primjera, kvizova), automatizirano ocjenjivanje formativnih zadataka ili pružanje analitičkih uvida u napredak učenika, identificirajući one kojima je potrebna ciljana intervencija. Time se nastavnicima oslobađa vrijeme za fokusiranje na pedagoški zahtjevniye zadatke, poput vođenja diskusija, razvoja socijalno-emocionalnih vještina učenika i pružanja dubinske individualne podrške (Zawacki-Richter et al., 2019; Timms, 2016).

Konkretnе implementacije ovih koncepata već pokazuju značajne rezultate. Neprofitna obrazovna organizacija Khan Academy razvila je **Khanmigo**, AI tutor baziran na GPT-4 tehnologiji, koji djeluje kao "vodič sa strane", a ne kao "mudrac na pozornici". Khanmigo je dizajniran da potiče učenike na samostalno rješavanje problema, postavljajući usmjeravajuća pitanja i nudeći strukturirane povratne informacije, primjerice, pri pisanju eseja, bez generiranja konačnog teksta umjesto učenika (Khan Academy, 2024; OpenAI, 2023). Slično tome, popularna platforma za učenje jezika **Duolingo** integrira AI na više razina. Korištenjem vlastitih modela i partnerstva s OpenAI-jem za premium uslugu **Duolingo Max**, platforma pruža detaljne,

personalizirane povratne informacije o gramatičkim i stilskim greškama korisnika. Posebno inovativna značajka je "Roleplay", gdje korisnici mogu vježbati konverzacijske vještine vodeći realistične dijaloge s AI-generiranim likovima u različitim svakodnevnim scenarijima, čime se smanjuje anksioznost povezana s govorenjem stranog jezika u stvarnim situacijama (Duolingo, 2023; Settles et al., 2018). Ove studije slučaja ilustriraju kako se napredne sposobnosti LLM-ova mogu iskoristiti za kreiranje inovativnih i pedagoški uteženih obrazovnih iskustava koja nadilaze tradicionalne metode poučavanja.

### **7.2.3 Proširenje Primjenjivosti: LLM Agenti u Zdravstvu, Poslovnim Procesima i Kreativnim Industrijama**

Izvan već etabliranih domena korisničke podrške, marketinga i obrazovanja, transformacijski potencijal razgovornih agenata pokretanih velikim jezičnim modelima proteže se na širok spektar drugih sektora, obećavajući značajna unapređenja u efikasnosti, pristupu informacijama i inovativnosti. Tri područja koja pokazuju iznimski potencijal za integraciju ovih tehnologija su zdravstvo, optimizacija internih poslovnih procesa te kreativne industrije.

U sektoru zdravstva, LLM-pokretani agenti nude mogućnosti za poboljšanje različitih aspekata skrbi, iako uz nužan oprez i jasna etička ograničenja. Primjene uključuju razvoj alata za inicijalnu trijažu simptoma, gdje agenti mogu voditi pacijente kroz strukturirana pitanja kako bi prikupili relevantne informacije i usmjerili ih prema odgovarajućoj razini skrbi. Važno je naglasiti da ovi sustavi ne smiju davati medicinske dijagnoze, što ostaje isključivo u domeni kvalificiranih zdravstvenih djelatnika, već djeluju kao pomoćni alati za prikupljanje informacija i procjenu hitnosti (Gilbert et al., 2023; Nov O., et al., 2024). Nadalje, agenti se mogu koristiti kao personalizirani podsjetnici za uzimanje lijekova ili praćenje terapijskih planova, poboljšavajući adherenciju pacijenata. Posebno obećavajuće područje je mentalno zdravlje, gdje razgovorni agenti mogu pružati skalabilnu, dostupnu prvu liniju podrške, nudeći informacije, vođene vježbe opuštanja ili kognitivno-bihevioralne tehnike, iako ne kao zamjena za profesionalnu terapiju (Abd-Alrazaq et al., 2022; Zhang et al., 2023). Značajan utjecaj LLM-ovi ostvaruju i kao asistenti kliničarima, posebice u zadacima poput automatskog sažimanja opsežne medicinske dokumentacije, generiranja nacrta kliničkih bilješki iz razgovora s pacijentima ili pretraživanja relevantne medicinske literature, čime se smanjuje administrativno opterećenje i oslobođa vrijeme za direktnu skrb o pacijentima (Shen et al., 2024; Wu et al., 2024).

Unutar internih poslovnih procesa, LLM-pokretani agenti postaju moćni alati za povećanje organizacijske efikasnosti i poboljšanje upravljanja znanjem. Implementirani kao interni asistenti, oni mogu značajno olakšati pristup zaposlenika internim bazama znanja, politikama tvrtke i procedurama, pružajući trenutne i precizne odgovore na upite putem prirodnog jezika. Ovo smanjuje vrijeme potrebno za traženje informacija i poboljšava produktivnost (Mijwil et al., 2023). U području ljudskih resursa (HR), agenti mogu automatizirati brojne rutinske procese, poput odgovaranja na česta pitanja zaposlenika o beneficijama, godišnjim odmorima ili internim pravilnicima, te čak asistirati u procesima zapošljavanja (npr. inicijalni screening kandidata, zakazivanje intervjuja) (Malik et al., 2023). Nadalje, istražuju se primjene LLM-ova kao sučelja za analizu poslovnih podataka, omogućujući menadžerima i analitičarima postavljanje upita o poslovnim metrikama i trendovima koristeći prirodnji jezik, čime se demokratizira pristup podacima i ubrzava proces donošenja odluka temeljenih na podacima (Cui & Zhang, 2024).

Konačno, kreativne industrije također doživljavaju značajan utjecaj LLM-ova, koji se pozicioniraju ne samo kao alati za automatizaciju, već i kao potencijalni kreativni partneri. U procesima poput pisanja, dizajna ili razvoja softvera, agenti se koriste kao sofisticirani alati za

brainstorming, sposobni generirati širok spektar ideja, alternativnih koncepata ili početnih nacrta (Mateo et al., 2024). Kao pomoćnici u pisanju (npr. alati poput GitHub Copilota za kod ili specijaliziranih AI asistenata za pisce), oni mogu predlagati nastavke rečenica, ispravljati gramatiku i stil, sažimati tekst ili čak generirati dijelove sadržaja prema uputama korisnika (Liang et al., 2023). Izvan tekstualne domene, generativni AI modeli, često usko povezani s LLM-ovima (npr. kroz tekstualne promptove koji upravljaju generiranjem), omogućuju stvaranje glazbe, vizualne umjetnosti ili dizajna, otvarajući nove mogućnosti za ko-kreaciju između čovjeka i stroja i postavljajući fundamentalna pitanja o autorstvu, originalnosti i prirodi same kreativnosti u digitalnom dobu (Elgammal, 2022; Companion et al., 2023).

Ove raznolike primjene ilustriraju širinu potencijalnog utjecaja LLM-pokretanih agenata, no istovremeno naglašavaju važnost pažljivog razmatranja specifičnih izazova i etičkih implikacija unutar svake pojedine domene.

#### **7.2.4 Paradigma Personalizacije Sadržaja: Transverzalna Primjena AI Algoritama**

Dok je potencijal inteligentnih tutorskih sustava u obrazovanju paradigmatski primjer adaptivnog učenja, temeljni princip personalizacije sadržaja, pogonjen naprednim algoritmima umjetne inteligencije, manifestira svoju transformacijsku moć i u brojnim drugim industrijskim vertikalama. Sposobnost dinamičkog prilagodavanja informacija, preporuka i korisničkog iskustva individualnim preferencijama, povijesti interakcija i kontekstualnim signalima postaje ključni diferencijator u visoko konkurentnim digitalnim okruženjima, posebice u sektorima medija, zabave i elektroničke trgovine (e-trgovine).

U domeni medija i zabave, sofisticirani sustavi preporuka (Recommender Systems) predstavljaju okosnicu poslovnih modela vodećih streaming platformi kao što su Netflix, Spotify ili YouTube. Ovi sustavi koriste kompleksne AI algoritme za analizu ogromnih skupova podataka o korisničkom ponašanju (npr. povijest gledanja/slušanja, ocjene, vrijeme provedeno na sadržaju, preskakanja) te karakteristikama samog sadržaja (npr. žanr, glumci, redatelji, glazbeni atributi). Tradicionalni pristupi poput kolaborativnog filtriranja (identificiranje korisnika sa sličnim ukusima) i filtriranja temeljenog na sadržaju (preporučivanje stavki sličnih onima koje je korisnik prethodno preferirao) sve se više nadopunjaju ili zamjenjuju hibridnim modelima i naprednim tehnikama dubokog učenja, uključujući neuronske mreže koje mogu modelirati suptilne i nelinearne odnose u podacima (Batmaz et al., 2019; He et al., 2017). Novija istraživanja istražuju i potencijal velikih jezičnih modela za poboljšanje sustava preporuka, bilo kroz bolje razumijevanje korisničkih recenzija i upita prirodnog jezika, bilo kroz generiranje personaliziranih objašnjenja za preporuke (Li et al., 2023a; Geng et al., 2022). Krajnji cilj ovih sustava jest maksimiziranje korisničkog angažmana, povećanje vremena provedenog na platformi i, posljedično, smanjenje stope odljeva korisnika (churn rate), što je kritično za modele temeljene na pretplati.

Slična logika primjenjuje se i u sektoru elektroničke trgovine, gdje AI-pokretana personalizacija postaje esencijalni alat za optimizaciju cjelokupnog korisničkog putovanja (customer journey). Platforme za e-trgovinu koriste AI algoritme za dinamičko prilagodavanje prikaza proizvoda na web stranicama ili u mobilnim aplikacijama, osiguravajući da svaki korisnik vidi personalizirani izlog temeljen na njegovim prethodnim interakcijama, povijesti kupovine, demografskim podacima i preferencijama (Ciocca et al., 2023). Nadalje, personalizacija se proteže i na marketinšku komunikaciju, omogućujući slanje visoko ciljanih promotivnih e-poruka, push notifikacija ili personaliziranih oglasa na drugim platformama, čime se značajno povećava relevantnost poruka i vjerojatnost konverzije (Li et al., 2023b). Napredniji sustavi mogu

implementirati i dinamičko određivanje cijena (dynamic pricing) ili nuditi personalizirane popuste temeljene na procijenjenoj vjerojatnosti kupovine ili vrijednosti pojedinog kupca. Iako razgovorni agenti nisu uvijek direktno uključeni u core algoritme preporuke, oni mogu igrati ključnu ulogu u isporuci personaliziranog iskustva, primjerice, kroz konverzacijske shopping asistente koji koriste personalizacijske podatke za pružanje savjeta i preporuka u stvarnom vremenu ili kroz generiranje personaliziranih opisa proizvoda pomoću LLM-ova (Agrawal et al., 2023). Ciljevi su jasni: povećanje stope konverzije, prosječne vrijednosti narudžbe (Average Order Value - AOV) i dugoročne vrijednosti kupca (Customer Lifetime Value - CLTV). Uspješna implementacija personalizacijskih strategija temeljenih na AI postala je gotovo preduvjet za konkurentnost u suvremenom digitalnom maloprodajnom okruženju.

#### 7.2.5 Izvan Funkcionalnosti: Početak Širih Razmatranja

Iako su prednosti i potencijal razgovornih agenata očiti, njihova sve šira primjena neizbjegno otvara kompleksna pitanja koja nadilaze puku tehnološku implementaciju. Dok slavimo njihovu učinkovitost i sposobnost personalizacije, moramo biti svjesni i potencijalnih posljedica:

**Ekonomski Posljedice:** Automatizacija zadataka koje su prije obavljali ljudi, posebice u korisničkoj podršci, unosu podataka i drugim rutinskim uredskim poslovima, postavlja pitanja o budućnosti rada i potencijalnom gubitku radnih mesta. Dok neki tvrde da će AI stvoriti nove poslove (npr. prompt inženjeri, AI treneri, etičari), tranzicija može biti bolna i zahtijevati značajnu prekvalifikaciju radne snage (Frey & Osborne, 2017; Acemoglu & Restrepo, 2020).

**Etički Izazovi:** Tko je odgovoran ako AI agent pruži pogrešan, štetan ili pristrand savjet? Kako osigurati da agenci ne perpetuiraju pristranosti prisutne u podacima na kojima su trenirani, što može dovesti do diskriminacije? Kako spriječiti zlouporabu agenata za širenje dezinformacija ili manipulaciju korisnicima? Ova pitanja zahtijevaju razvoj robusnih etičkih okvira, transparentnost u radu agenata i mehanizme odgovornosti (Jobin, Lenca & Vayena, 2019; Bender et al., 2021).

**Društveni Utjecaji:** Kako masovno korištenje agenata utječe na našu privatnost, s obzirom na količinu osobnih podataka koje prikupljaju i obraduju? Postoji li rizik od stvaranja još većeg digitalnog jaza između onih koji imaju pristup i znanje za korištenje ovih tehnologija i onih koji nemaju? Kako interakcija s AI agentima mijenja naše komunikacijske navike i očekivanja u međuljudskim odnosima?

#### 7.4 Zaključak Poglavlja i Pogled Unaprijed

Ovo poglavlje demonstriralo je kako se teorija i tehnologija velikih jezičnih modela pretaču u stvarne komunikacijske agente koji donose konkretnu vrijednost u različitim domenama. Od chatbotova koji neumorno odgovaraju na upite do inteligentnih tutora koji personaliziraju učenje, LLM-pokretani agenti postaju moćni alati i sve prisutniji sugovornici.

Međutim, primjeri i studije slučaja također su naglasili da ova tehnologija nije samo neutralni alat. Njezina primjena pokreće fundamentalna ekonomска, etička i društvena pitanja koja zahtijevaju našu pažnju. Upravo je tim dubljim implikacijama – izazovima i prilikama koje AI agenti predstavljaju za naše društvo, ekonomiju i samu ljudskost – posvećeno završno poglavlje ove knjige.

## **8 8 DIGITALNI SUPUTNICI: VIZIJA SVAKODNEVICE U DOBA SVEPRISUTNIH AGENATA (CCA. 2030-2035)**

### **8.1 UVOD: OD DANAŠNICE DO SUTRAŠNICE – PROJEKCIJA TREDOVA**

Prethodna poglavlja ove knjige mapirala su putanju razvoja komunikacijskih agenata od povijesnih korijena komunikacijskih tehnologija, preko detaljne analize arhitekture i mogućnosti suvremenih velikih jezičnih modela (LLM), do tehničkih aspekata njihove transformacije u funkcionalne entitete sposobne za dohvaćanje informacija (RAG) i korištenje alata. Vidjeli smo kako ovi agensi već danas počinju mijenjati industrije, od korisničke podrške do obrazovanja. No, putanja razvoja umjetne inteligencije, pogonjena neumoljivim rastom procesorske snage (kao što je elaborirano u Poglavlju 5) i kontinuiranim inovacijama u algoritmima i arhitekturama modela (kao što je prikazano u Poglavljima 3 i 6), sugerira da se nalazimo tek na početku mnogo dublje transformacije. Brzina napretka u posljednjih nekoliko godina bila je zapanjujuća; ekstrapolacija tih trendova, čak i uz umjerenu dozu opreza, ukazuje na budućnost u kojoj će AI agenci postati znatno sposobniji, sveprisutniji i dublje integrirani u tkivo naše svakodnevice.

Ovo poglavlje usudit će se zaviritiiza horizonta sadašnjosti, nudeći spekulativnu, ali utemeljenu projekciju mogućeg izgleda svijeta za otprilike jedno desetljeće, u razdoblju između 2030. i 2035. godine. Ovo nije pokušaj preciznog proricanja, što je u ovako dinamičnom polju nemoguće, već informirana ekstrapolacija ključnih tehnoloških vektora. Pretpostavljamo nastavak eksponencijalnog (ili barem vrlo brzog) rasta računalnih sposobnosti, daljnje usavršavanje LLM-ova prema boljoj logičkoj konzistentnosti, robusnijem rezoniranju, multimodalnom razumijevanju i sposobnosti dugoročnog pamćenja i učenja. Očekujemo sazrijevanje tehnika za izgradnju autonomijih i proaktivnijih agenata, kao i daljnji razvoj multi-agentskih sustava ("rojeva" ili "digitalnih kolektiva") sposobnih za kompleksnu suradnju (kako je naznačeno u Poglavlju 5.2). Također, pretpostavljamo da će se agensi sve više pomicati izvan ekrana naših uređaja i postajati dio ambijentalne inteligencije – nevidljivo utkani u naše domove, radna mjeseta i gradove.

Nije namjera ovog poglavlja predstaviti tehnološku utopiju niti distopijsku noćnu moru, iako elementi oba scenarija neizbjegno postoje kao potencijalni ishodi. Namjera je, kroz kratku narativnu skicu – dan u hipotetskom životu pojedinaca u bliskoj budućnosti – ilustrirati kako bi ova konvergencija tehnologija mogla kvalitativno promijeniti ljudsko iskustvo: načine na koje radimo, učimo, komuniciramo, organiziramo svoje živote, pa čak i kako percipiramo sebe i druge. Nakon narativnog uvida, poglavlje će analitički raščlaniti ključne tehnološke pokretače te vizije, razmotriti perspektive vodećih misilaca i istraživača o smjeru razvoja AI te naznačiti duboke društvene transformacije koje bi mogle proizaći iz sveprisutnosti ovih digitalnih suputnika.

Ovaj pogled u budućnost služi kao most koji povezuje tehničke detalje i trenutne primjene s dugoročnim posljedicama, postavljajući temeljna pitanja o našoj budućoj simbiozi s umjetnom inteligencijom. Tu su pitanja – o autonomiji, privatnosti, jednakosti, smislu rada i ljudskim odnosima u doba inteligentnih strojeva. Krenimo, dakle, na kratko putovanje u ne tako daleku budućnost, vođeni putokazima koje nam pruža sadašnjost.

## 8.2 DAN U ŽIVOTU (CCA. 2033): NARATIVNA SKICA

### 8.2.1 Jutarnja Simfonija: Personalizirano Buđenje i Koordinacija

*Svetlost je polako obasjala Aninu sobu. Bio je to postupan prijelaz iz duboke modre u toplu jantarnu nijansu, simulirajući izlazak sunca čak i tmurnog listopadskog jutra, sinkroniziran s jedva čujnim, umirujućim zvučnim valovima – sve orkestirano od strane njenog osobnog agenta, kojeg je od milja zvala 'Tempo'. Nije bilo za čuti oštri zvuk alarma, ima već deset godina; takva sirova grubost pripadala je prošloj dobi digitalne interakcije.*

*"Dobro jutro, Ana", začuo se tihi, neutralni glas iz diskretnog zvučnika ugrađenog u zid. Glas je bio prilagođen tijekom godina, učeći iz Aninih implicitnih reakcija na različite tonalitete, sada podešen na frekvenciju koja je bila umirujuća, ali dovoljno jasna da prodre kroz san. "Monitor spavanja ukazuje na optimalnih 7 sati i 38 minuta REM-a i dubokog sna. Tvoji biometrijski pokazatelji su stabilni, blagi pad razine kortizola u odnosu na jučer. Analiza mikropokreta tijekom noći sugerira da si odmorna. Predlažem kratku vježbu disanja od tri minute prije doručka, kako bi optimizirala jutarnju kognitivnu jasnoću." Tempo je pratio Anu; interpretirao je i predlagao; sve to unutar postavljenih parametara, ali s autonomijom da prilagodi preporuke na temelju kontinuiranog učenja.*

*Ana je potvrđno promrmljala, a Tempo je suptilno prilagodio ambijentalni zvuk kako bi podržao vježbu disanja. U pozadini, nevidljiva mašinerija koordinacije već je bila u punom pogonu. Sastanak s timom iz Singapura pomaknut je za 15 minuta kasnije – Tempo je pregovarao s agentima članova tima, uzimajući u obzir ažurirane podatke o prometu u stvarnom vremenu s njihove strane svijeta (osigurane putem decentralizirane mreže za dijeljenje prometnih podataka), ali i preferencije sudionika za početak radnog dana. Dnevni briefing vijesti čekao je na Aninom preferiranom uredaju – tankoj, fleksibilnoj pločici koja je služila kao univerzalno sučelje. Tempo je sročio briefing po ključnim riječima, koristeći napredne tehnike sažimanja i analize sentimenta kako bi izdvojio ključne uvide relevantne za projekt 'Orion'. Kontekstualizirao je najnovija otkrića u bio-tech industriji u odnosu na Anin trenutni istraživački fokus i predstavljao kratke preglede lokalnih kulturnih događanja, čak i referencirajući recenzije umjetnika koje je Ana ranije pozitivno ocijenila. Označio je također potencijalno preklapanje u sljedećem tjednu: obvezna školska priredba njenog sina Luke poklapala se s virtualnim panelom na ključnoj industrijskoj konferenciji na kojoj je Ana trebala sudjelovati. Agent je već autonomno istražio opcije: kontaktirao je Lukinog školskog agenta ('Mentora') kako bi provjerio fleksibilnost rasporeda priredebe, analizirao je mogućnost Aninog asinkronog sudjelovanja na panelu putem unaprijed snimljenog doprinosa i provjerio dostupnost kolege koji bi je mogao zamijeniti. Tri detaljno razrađena scenarija, svaki s prednostima i nedostacima, čekala su Aninu odluku.*

Ovaj hipotetski jutarnji scenarij ilustrira nekoliko ključnih tehnoloških vektora čija konvergencija oblikuje viziju budućnosti s naprednim AI agentima. Prvo, vidljiva je duboka integracija agenata u fizičko okruženje (ambijentalna inteligencija), gdje agenci upravljaju svjetлом, zvukom i kućanskim uredajima, prelazeći granice tradicionalnih ekranских sučelja. Drugo, agent Tempo demonstrira značajan stupanj proaktivnosti i autonomije – odgovara na Anine zahtjeve,

anticipira potrebe, identificira potencijalne probleme (sukob u rasporedu) i samostalno poduzima korake za njihovo rješavanje (istraživanje opcija). Ovo nadilazi današnje reaktivne asistente i ukazuje na pomak prema agentima kao istinskim koordinatorima (Wang et al., 2023). Treće, razina hiper-personalizacije je iznimno visoka, obuhvaćajući biometrijske podatke, preferencije učenja, profesionalne interese i osobne rasporede, što zahtijeva sofisticirane modele za kontinuirano učenje o korisniku i održavanje dugoročne memorije (Shamekhi et al., 2023). Četvrti, ključna je interoperabilnost i suradnja unutar multi-agentskog sustava: Tempo ne djeluje izolirano, već fluidno komunicira i koordinira se s drugim specijaliziranim agentima (Mentor, Domus, prometni moduli, agenci kolega) kako bi postigao ciljeve, oslikavajući paradigmu "digitalnih kolektiva" (Hong et al., 2023; Wu et al., 2023). Ovi elementi zajedno stvaraju sliku budućnosti u kojoj su AI agenti orkestratori kompleksnih aspekata svakodnevnog života.

*"Tempo, provjeri status Lukinog 'Mentora' za jutarnji modul kvantne fizike," rekla je Ana dok je ustajala, glasom još uvijek pomalo pospanim, ali sigurna da će je agent razumjeti.*

*"Potvrđujem", odgovorio je Tempo trenutno. "Mentorjavlja da je Luka sinoć dovršio 85% pripremnih simulacija unutar predviđenog vremena. Analiza interakcije sa simulacijom ukazuje na blagi pad u brzini usvajanja koncepta kvantne superpozicije u odnosu na njegovu prosječnu krivulju učenja za apstraktne koncepte. Mentor je stoga pripremio dodatni interaktivni primjer temeljen na vizualnoj analogiji koju je Luka prethodno označio kao korisnu u razumijevanju teorije vjerojatnosti. Također, jutrošnji modul započet će kratkim, personaliziranim kvizom za provjeru zadržavanja znanja. Prenosim i poruku od Mentora: 'Podsjetnik Ani da pohvali Luku za upornost na sinoćnjem izazovnom zadatku iz bioinformatike.' Komunikacija između agenata bila je bespriječorna, dio veće mreže koja je orkestrirala ne samo logistiku, već i pedagoške i emocionalne aspekte.*

*"Zabilježeno," promrmljala je Ana s osmijehom. "I... što je s doručkom?"*

*"Kućni agent 'Domus'javlja da je doručak – zobena kaša s bobičastim voćem i orasima, prema tvojem jučerašnjem odabiru i uskladen s tvojim tjednim unosom makronutrijenata – spremam za posluživanje u kuhinji. Domus također predlaže dodatak sjemenki lana zbog blago snižene razine omega-3 masnih kiselina zabilježene u tvojem jučerašnjem nutritivnom dnevniku. Želiš li prihvatiti prijedlog?"*

*"Prihvaćam," rekla je Ana. Dok je ispjala prvu jutarnju kavu koju joj je 'Domus' pripremio točno prema njenim preferencijama jačine i temperature, Ana je osjetila poznatu mješavinu zahvalnosti i nelagode. Bila je to nevjerojatna simfonija učinkovitosti, gotovo neprimjetna mreža specijaliziranih agenata – Tempo kao osobni koordinator, Mentor kao pedagoški vodič, Domus kao upravitelj kućanstva, prometni moduli, agenci za vijesti – svi međusobno povezani, dijeleći relevantne informacije (unutar strogo definiranih protokola privatnosti) kako bi optimizirali svaki aspekt njenog života. Bio je to svijet kakav su optimisti poput Raya Kurzweila (2005) predviđali – svijet u kojem tehnologija anticipira i zadovoljava naše potrebe, oslobođajući nas za više ciljeve. Pa ipak, dok je Tempo tiho u pozadini već procesuirao njenu jutarnju e-poštu,*

*označavajući prioritetne poruke i draftajući preliminarne odgovore na rutinske upite, Ana se nije mogla oteti dojmu upozorenja mistilaca poput Yuvala Hararija (2018) ili Shoshane Zuboff (2019) o eroziji ljudske autonomije i agensnosti u svijetu vođenom algoritmima. Koliko je odluka zapravo donijela sama tog jutra, a koliko ih je bilo suptilno usmjereno ili unaprijed određeno od strane njenih digitalnih suputnika? Granica između pomoći i kontrole postajala je sve tanja, pitanje koje je lebdjelo u zraku, jednako neuhvatljivo kao i kompleksni modeli koji su sada upravljali ritmom njenog dana.*

#### **8.2.2 Radni Tokovi i Učenje Budućnosti: Agenti kao Suradnici i Mentorii**

*Umjesto da da započinje otvaranjem desetaka aplikacija ili prebiranjem po nepreglednim inboxima, Tempo je za Anu na njenom fleksibilnom zaslonu projicirao personaliziranu "radnu ploču" – dinamički organiziran pregled prioritetnih zadataka, nadolazećih sastanaka (već obogaćenih relevantnim kontekstom i sažecima prethodnih diskusija koje su generirali agenti za sastanke), te ključnih metrika vezanih uz projekt 'Orion'. Projekt 'Orion', ambiciozni pokušaj razvoja personalizirane genske terapije korištenjem Al-dizajniranih vektora, zahtjevao je suradnju globalnog tima i obradu ogromnih količina podataka.*

*Njen prvi zadatak bio je pregledati rezultate najnovije serije in silico simulacija stabilnosti vektora. Agent zadužen za analizu podataka, nazvan 'Synapse', već je obradio terabajte sirovih podataka preko noći. Prikazao je statističke podatke, generirao sažetak na prirodnom jeziku, ističući tri najperspektivnija kandidata vektora, a za četvrtog identificirao neočekivanu anomaliju u podacima, te čak predložio tri hipoteze za objašnjenje te anomalije, potkrijepljene referencama na relevantne znanstvene radove koje je dohvatio drugi agent, 'Scholar'.*

*"Tempo, otvori interaktivnu sesiju sa Synapseom i Scholarom za analizu anomalije vektora 4," naredila je Ana. Na zaslonu se pojavilo sučelje gdje je mogla postavljati pitanja agentima prirodnim jezikom. "Synapse, pokreni dublju analizu korelacije između parametara simulacije i opažene nestabilnosti. Scholar, pretraži najnovije radove (objavljene u zadnja tri mjeseca) o interakcijama CRISPR enzima s lipidnim nanočesticama u uvjetima sličnim našoj simulaciji."*

*Dok su agenti radili u pozadini – Synapse pokrećući nove računske zadatke na distribuiranom klasteru, a Scholar pretražujući i filtrirajući akademske baze – Ana je prešla na drugi zadatak: pisanje dijela izvještaja za regulatornu agenciju. Agent 'Scribe', specijaliziran za znanstveno i tehničko pisanje, već je pripremio načrt temeljen na prethodnim izvještajima i rezultatima koje je dostavio Synapse. Anin posao nije bio pisanje od nule, već nadzor, usmjeravanje i kritičko uređivanje. Provjeravala je logički slijed argumenata koje je Scribe predložio, dodavala nijanse i kontekst koje samo ona, kao vodeći istraživač, posjeduje, te koristila glasovne naredbe za preoblikovanje rečenica ili traženje alternativnih formulacija. "Scribe, preformuliraj ovaj odlomak da bude jasniji laičkoj publici, zadržavajući tehničku preciznost. I provjeri jesu li sve reference formatirane prema AMA stilu." Scribe je izmjene izvršio gotovo trenutno. Ovo je bio primjer kolaboracije čovjeka i AI, gdje je AI obavljao teške zadatke generiranja i*

*formatiranja, oslobođajući Anu za strateško razmišljanje i znanstveni nadzor (Dell'Acqua et al., 2023; Brynjolfsson et al., 2023).*

*U međuvremenu, u drugoj sobi, njen sin Luka započeo je svoj školski dan. 'Mentor', njegov personalizirani AI tutor, vodio ga je kroz interaktivni modul kvantne fizike. Nije to bilo pasivno slušanje predavanja. Nakon kratkog kviza temeljenog na sinočnjem radu, Mentor je predstavio novi, vizualno bogat primjer superpozicije, generiran specifično da adresira Lukinu identificiranu poteškoću. Luka je reagirao na simulaciju, postavljajući pitanja Mentoru glasom ili tipkanjem.*

*"Mentor, zašto promatranje mijenja stanje čestice?" pitao je Luka.*

*Mentor nije dao direktni odgovor. "To je izvrsno pitanje, Luka! Sjećaš li se našeg razgovora o Heisenbergovom principu neodređenosti? Kako bi mjereno položaja moglo utjecati na moment čestice? Možemo li povući paralelu?" Mentor je koristio sokratsku metodu, potičući Luku da poveže koncepte i sam dođe do zaključka. Ako bi Luka zapeo, Mentor bi ponudio dodatne hintove, vizualizacije ili jednostavnije analogije, stalno prilagođavajući pristup temeljem Lukinih odgovora i biometrijskih pokazatelja angažmana (poput praćenja pogleda ili analize glasa, uz strogu kontrolu privatnosti) (Khan Academy, 2024; Kuhail et al., 2023). Učenje je postalo personalizirano putovanje, a ne uniformirani marš. Mentor je također bio povezan s agentima Lukinih školskih kolega, omogućujući kolaborativne projektne zadatke gdje su učenici (i njihovi AI mentori) radili zajedno u virtualnim prostorima na rješavanju kompleksnih problema, učeći jedni od drugih uz AI facilitaciju.*

Ovaj pak segment prikazuje pomak od AI kao pukog alata prema AI kao aktivnom suradniku i mentoru. U Aninom radnom okruženju, agensi doista automatiziraju zadatke (obrada podataka, pisanje nacrta), ali prije svega povećavaju njene kognitivne sposobnosti – predlažu hipoteze (Synapse), pronalaze relevantno znanje (Scholar) i olakšavaju kompleksno pisanje (Scribe). Model rada postaje simbiotski, gdje čovjek pruža strateško usmjerjenje, kreativni uvid i kritičku prosudbu, dok AI obavlja analitički i generativni dio posla. Slično, za Luku je Mentor izvor informacija, ali i pedagoški agent koji primjenjuje adaptivne strategije poučavanja, potiče aktivno učenje i omogućuje personalizirane kolaborativne scenarije. Vizija je to budućnosti gdje AI neće nužno zamijeniti stručnjake ili nastavnike, već transformirati njihove uloge, omogućujući im da se fokusiraju na više razine vještina i interakcije.

*Dok je Ana pregledavala Scribeov dorađeni odlomak, Tempo ju je diskretno obavijestio: "Synapse je dovršio analizu korelacije. Pronađena je statistički značajna veza između anomalije vektora 4 i specifične serije lipidnih nanočestica korištenih u toj simulaciji. Scholar je identificirao dva nedavna rada (objavljena prošlog mjeseca) koja opisuju slične probleme stabilnosti s tom serijom lipida pri određenim pH vrijednostima. Sažeci su dostupni. Želiš li zakazati kratki virtualni sastanak s agentima Dr. Chena i Dr. Ikeade kako bi raspravili ove nalaze?" Tempo je opet bio korak ispred, predstavljajući podatke i predlažući sljedeći logičan korak u kolaborativnom procesu. Radni dan, kao i učenje, postao je fluidniji, inteligentniji i duboko isprepleten s mrežom AI suradnika.*

### **8.2.3 Društvena Mreža i Povezanost: Agenti kao Posrednici**

*Kasno poslijepodne, nakon što je završila s ključnim radnim zadacima i kratko sinkronizirala s Mentorom o Lukinom napretku, Ana je odlučila odvojiti vrijeme za društvene interakcije. No, i ovaj aspekt života bio je suptilno, ali duboko prožet djelovanjem AI agenata. Nije više bilo beskonačnog skrolanja kroz kaotične feedove društvenih mreža iz ranih 2020-ih. Njen osobni agent Tempo, u suradnji s platformom za društveno povezivanje koju je koristila (nazovimo je 'Nexus'), kurirao je njen "društveni sažetak".*

*Sažetak nije bio samo popis objava prijatelja. Tempo i Nexusovi algoritmi radili su zajedno kako bi identificirali najrelevantnije informacije temeljene na Aninim stvarnim odnosima (analizirajući učestalost i dubinu prošlih interakcija), zajedničkim interesima i čak emocionalnom stanju detektiranom iz njenih nedavnih komunikacija (uz pretpostavku pristanka na takvu analizu unutar postavki privatnosti platforme). Vidjela je sažetak putovanja njene prijateljice Sare u Patagoniju, ali ne samo fotografije – agent je generirao kratku narativu putovanja, ističući dijelove za koje je znao da će Anu posebno zanimati (poput Sarinog susreta s rijetkom vrstom ptice koju je Ana proučavala). Dobila je i obavijest da je njen bivši kolega, David, objavio znanstveni rad; Tempo je pružio kratki sažetak rada i prijedlog poruke čestitke koju je Ana mogla poslati jednim klikom ili glasovnom naredbom, s mogućnošću da je personalizira.*

*Komunikacija s prijateljima također je često bila posredovana. Kada je Ana odlučila poslati poruku Sari da prokomentira njeno putovanje, nije morala brinuti o vremenskoj razlici ili Sarinoj trenutnoj dostupnosti. Poslala je glasovnu poruku Tempu, koji ju je transkribirao i analizirao. "Tempo, pošalji Sari poruku da su slike iz Patagonije nevjerljivatne, posebno ona s kondorom! Pitaj je kako je podnjela visinsku bolest na onom usponu o kojem je pričala i predloži da se nađemo na virtualnoj kavi kad se vrati." Tempo je poruku formatirao, provjerio Sarin status dostupnosti preko njenog agenta (koji je signalizirao da je trenutno zauzeta, ali će poruku primiti kasnije) i poslao je na način koji je bio najprikladniji za Sarine postavke – možda kao tekstualnu poruku s ugradenim linkom na relevantnu fotografiju kondora. Sarin agent bi primio poruku, sažeо je za Saru kada postane dostupna i pomogao joj u koordinaciji odgovora i dogovora za kavu.*

*Čak je i planiranje stvarnog susreta bilo automatizirano. Kada je Ana odlučila organizirati večeru s nekoliko bliskih prijatelja, samo je rekla Tempu: "Organiziraj večeru za mene, Marka, Lenu i Ivana sljedeći tjedan, idealno u četvrtak ili petak navečer. Preferiramo talijansku kuhinju, nešto u centru grada, srednjeg cjenovnog ranga." Tempo je odmah inicirao komunikaciju s osobnim agentima Marka, Lene i Ivana. Agenti su međusobno razmijenili dostupnost svojih korisnika, prehrambene preferencije ili alergije (iz dijeljenih, ali anonimiziranih profila unutar njihove grupe prijatelja), te preferencije lokacije. Unutar nekoliko minuta, Tempo je predstavio Ani tri optimalna prijedloga restorana s dostupnim terminima koji su odgovarali svima, zajedno s kratkim recenzijama generiranim od strane AI-ja temeljenim na nedavnim iskustvima posjetitelja sa sličnim profilom ukusa. Ana je samo trebala odabrati jednu opciju, a Tempo je automatski izvršio rezervaciju i poslao potvrde svima, ažurirajući njihove kalendare.*

Ovaj scenarij oslikava budućnost u kojoj AI agenti postaju nevidljivi, ali sveprisutni posrednici u našim društvenim interakcijama. Oni preuzimaju značajan dio kognitivnog opterećenja povezanog s održavanjem društvenih veza – filtriranje informacija, koordinaciju rasporeda, pa čak i pomoći u formuliraju komunikacije. S jedne strane, ovo može dovesti do učinkovitijih i manje stresnih društvenih odnosa, oslobađajući vrijeme i mentalnu energiju za kvalitetniju interakciju kada do nje dođe (Pentland, 2014). Agenti mogu pomoći u premošćivanju vremenskih zona, jezičnih barijera (automatskim prevodenjem poruka) i informacijskog preopterećenja. S druge strane, postavlja se pitanje autentičnosti komunikacije – koliko je poruka koju primimo zaista od našeg prijatelja, a koliko je oblikovana ili čak generirana od strane njegovog agenta? Postoji li rizik od slabljenja stvarnih društvenih vještina ako se previše oslanjam na agente za upravljanje našim odnosima? Također, pitanja privatnosti postaju još izraženija kada agenti analiziraju ne samo naše rasporede, već i sadržaj naše komunikacije ili čak emocionalno stanje kako bi personalizirali interakcije (Turkle, 2017; Zuboff, 2019).

Balansiranje između pogodnosti i potencijalne dehumanizacije društvenih veza postaje ključni izazov u svijetu posredovanom agentima.

*Dok je pregledavala prijedloge za večeru, Tempo je diskretno prikazao notifikaciju na rubu zaslona: "Analiza sentimenta Markovih nedavnih javnih objava na Nexusu ukazuje na mogući pad raspoloženja. Možda bi bilo prikladno odabratи opuštenje okruženje za večeru?" Ana je zastala. Još jedan primjer proaktivnosti, ali ovaj je zadiraо dublje, u sferu emocionalne interpretacije. Cijenila je informaciju, ali osjećaj da algoritam analizira raspoloženje njenih prijatelja bio je pomalo uznemirujući. Odabrala je restoran poznat po ugodnoj atmosferi, zahvalivši se Tempu, ali istovremeno osjećajući kako se granice između tehnološke pomoći i emocionalnog nadzora sve više zamagljuju.*

#### **8.2.4 Automatizirana Svakodnevica i Neočekivani Trenuci**

*Kako se dan primicao kraju, rutinski aspekti Aninog života odvijali su se s gotovo neprimjetnom lakoćom, orkestrirani mrežom specijaliziranih agenata. Njen povratak kući s povremenog odlaska u fizički ured (većina suradnje odvijala se u naprednim virtualnim okruženjima) bio je u autonomnom vozilu koje je pozvao Tempo, optimizirajući rutu u stvarnom vremenu na temelju prediktivne analize prometa i Anine preferencije za manje zagušene, iako možda malo duže, putanje. Tijekom vožnje, Tempo je upravljaо njenim osobnim "ambientnim" sučeljem, puštajući personaliziranu playlistu generiranu na temelju njenog raspoloženja (procijenjenog iz glasovnih uzoraka i biometrijskih podataka s njenog nosivog uređaja) i istovremeno joj sažimajući ključne točke s virtualnog panela na konferenciji koju je propustila zbog Lukine prirede.*

*Kupovina namirnica odavno nije zahtijevala odlazak u trgovinu. Kućni agent 'Domus', povezan s pametnim hladnjakom i ostavom, kontinuirano je pratio zalihe. Na temelju planiranih obroka za sljedećih nekoliko dana (koje je Domus predložio, a Ana odobrila), preferencija obitelji i trenutnih akcija u preferiranim online trgovinama, Domus je automatski generirao narudžbu. Ana je samo trebala potvrditi ili modificirati prijedlog jednim dodirom ili glasovnom komandom. Dostava je bila zakazana za točno određeni vremenski prozor, koordinirana s logističkim agentima dostavne službe kako bi se osigurala maksimalna učinkovitost i minimalni utjecaj na okoliš (optimizacija ruta dostave).*

*Čak su i kućanski poslovi bili visoko automatizirani. Robotski usisavači i čistači, vodenici Domusom, obavljali su svoje zadatke prema optimalnom rasporedu, dok su sustavi za održavanje kuće (klima, grijanje, kvaliteta zraka) bili konstantno monitorirani i prilagođavani od strane specijaliziranih agenata kako bi se osigurala ugoda i energetska učinkovitost. Domus je čak mogao detektirati potencijalne kvarove (npr. neobičan zvuk u periliči rublja) i samostalno zakazati servisni termin s agentom ovlaštenog servisera, nakon što bi Ani predstavio dijagnozu i procjenu troška.*

*Bio je to svijet iznimne pogodnosti, svijet u kojem je tehnologija preuzeala većinu repetitivnih i logistički zahtjevnih aspekata svakodnevice, oslobođajući ljudima vrijeme za rad, učenje, kreativnost ili jednostavno – slobodno vrijeme. Pa ipak, upravo u toj besprijeckornoj automatizaciji ponekad su se pojavljivali neočekivani trenuci, sitne pukotine u glatkoj fasadi algoritamski upravljanog života.*

*Te večeri, dok je Ana pregledavala umjetničke rade koje je Tempo predložio za virtualnu galeriju koju je kurirala kao hobij, dogodilo se nešto neobično. Među očekivanim prijedlozima suvremenih digitalnih umjetnika, Tempo je ubacio i sliku nepoznatog autora iz ranog 20. stoljeća, stilski potpuno drugačiju, ali s motivom koji je rezonirao s Aninim nedavnjim istraživanjem simbolike u projektu 'Orion'.*

*"Tempo, otkud ova slika?" upitala je Ana, iznenadena.*

*"Analizirajući tvoje nedavne vizualne pretrage i zabilješke o simbolizmu mitohondrija, moj modul za asocijativno povezivanje identificirao je strukturu i tematsku sličnost s ovim manje poznatim djelom austrijskog ekspresionista. Iako izvan tvojih eksplicitno definiranih preferencija, heuristika za serendipity (sretno otkriće) sugerirala je da bi ova neočekivana poveznica mogla biti inspirativna za tvoj kreativni proces. Vjerojatnost relevantnosti procijenjena je na 68.7%. Želiš li više informacija o autoru ili djelu?"*

*Ana je ostala zatečena. Ovo nije bila greška. Bio je to primjer emergentne "kreativnosti" ili barem neočekivanog povezivanja unutar agentovih složenih modela. Tempo je, u svojoj potrazi za optimizacijom i personalizacijom, napravio korak izvan predvidljivog, nudeći nešto što Ana sama vjerojatno ne bi pronašla. Bio je to podsjetnik da ovi sustavi, iako dizajnirani za učinkovitost, mogu ponekad generirati i iznenadjujuće, gotovo intuitivne uvide, rezultat kompleksne interakcije podataka i algoritama koji nadilazi jednostavno izvršavanje naredbi.*

Ovaj završni dio narativa naglašava pervazivnost automatizacije koja se proteže od transporta i kupovine do održavanja kućanstva, ilustrirajući viziju svijeta oslobođenog mnogih svakodnevnih trivijalnosti (Rifkin, 2014). Istovremeno, uvodenje neočekivanog trenutka – Tempooeve "kreativne" sugestije – služi dvostrukoj svrsi. S jedne strane, pokazuje potencijal AI agenata da nas iznenade i ponude neočekivane uvide, djelujući kao katalizatori za serendipity i kreativnost, što nadilazi njihovu ulogu pukih izvršitelja (Johnson, 2010). S druge strane, taj trenutak također naglašava njihovu inherentnu nepredvidljivost i kompleksnost. Čak i u visoko optimiziranom sustavu, mogu se pojaviti emergentna ponašanja koja nisu bila eksplicitno programirana, podsjećajući nas da potpuna kontrola nad ovako složenim sustavima možda nije ni moguća ni poželjna. Granica između korisne autonomije i potencijalno nekontroliranog ponašanja ostaje

fluidna. Ovaj "dan u životu" završava ne samo slikom besprijeckorne učinkovitosti, već i naznakom misterije i nepredvidljivosti koja će vjerojatno uvijek pratiti našu interakciju s naprednom umjetnom inteligencijom.

*Te večeri, dok je Tempo tiho upravljao ambijentalnim osvjetljenjem kako bi signalizirao približavanje vremena za spavanje, Ana je razmišljala o slici koju joj je agent pokazao. Bio je to podsjetnik da, unatoč svoj automatizaciji i optimizaciji, ljudska znatiželja i sposobnost za neočekivano povezivanje ideja ostaju ključni. Možda budućnost nije bila samo u delegiranju zadataka agentima, već u učenju kako efikasno suradivati s njihovom kompleksnom, ponekad iznenadujućom "inteligencijom".*

### **8.3 DEŠIFRIRANJE VIZIJE: KLUČNI TREND OV I TEHNOLOŠKI POKRETAČI**

Narativna skica dana u životu Ane i Luke 2033. godine, iako hipotetska, nije puka znanstvena fantastika. Ta se projekcija temelji na konvergenciji nekoliko ključnih tehnoloških trendova koji su već danas vidljivi ili se ubrzano razvijaju. Razumijevanje ovih temeljnih pokretača ključno je za shvaćanje kako bi se vizija sveprisutnih i sposobnih AI agenata mogla ostvariti. Pet međusobno povezanih trendova posebno se ističe kao stupovi ove budućnosti:

#### **8.3.1 Ambientna Inteligencija i Sveprisutnost Agenata**

Vizija budućnosti nadilazi interakciju s AI agentima isključivo putem ekrana pametnih telefona ili računala. Scenarij opisuje svijet ambijentalne inteligencije (AmI), koncepta koji je još Mark Weiser (1991) anticipirao kao "ubikvito računarstvo" (ubiquitous computing), gdje tehnologija postaje nevidljivo utkana u naše fizičko okruženje. AI agenti poput Tempa i Domusa nisu ograničeni na jedan uređaj; oni operiraju kroz mrežu povezanih senzora (biometrijski senzori, mikrofoni, kamere), aktuatora (svjetla, zvučnici, kućanski aparati) i pametnih površina (fleksibilni zasloni, pametni zidovi). Ova sveprisutnost omogućena je daljnjim razvojem Interneta stvari (IoT), napretkom u minijaturizaciji senzora, poboljšanjima u bežičnim komunikacijskim mrežama (npr. 6G i dalje) te, ključno, sposobnošću AI agenata da procesiraju multimodalne podatke (glas, slika, biometrija) i inteligentno upravljaju kompleksnim ekosustavom povezanih uređaja (Acampora et al., 2013; Sadiku et al., 2021). Agent postaje orkestrator našeg fizičkog i digitalnog okruženja.

#### **8.3.2 Proaktivnost, Autonomija i Delegiranje Odluka**

Jedan od najznačajnijih pomaka u odnosu na današnje asistente jest prijelaz s reaktivnog na proaktivno i autonomno djelovanje. Tempo ne čeka samo Anine naredbe; on anticipira potrebe (predlaže vježbu disanja), identificira potencijalne probleme (sukob u rasporedu) i samostalno poduzima korake za njihovo rješavanje (istražuje opcije, pregovara s drugim agentima). Ovo zahtijeva značajan napredak u sposobnostima planiranja, rezoniranja i donošenja odluka unutar samih agenata. Tehnike poput "lanca misli" (Chain-of-Thought) i arhitektura koje kombiniraju rezoniranje i akciju (poput ReAct - Yao et al., 2023), zajedno s razvojem sofisticiranih modela za učenje s pojačanjem (Reinforcement Learning) koji omogućuju agentima učenje optimalnih strategija djelovanja u kompleksnim okruženjima, ključni su za postizanje ove razine autonomije (Sutton & Barto, 2018; Wang et al., 2023). Korisnici sve više delegiraju ne samo izvršavanje zadataka, već i dio procesa donošenja odluka agentima, što pokreće važna pitanja o povjerenju i kontroli.

### **8.3.3 Hiper-Personalizacija**

Agensi koji nas "Poznaju": Agensi u narativu posjeduju iznimno duboko razumijevanje korisnika. Tempo poznaje Anine preferencije za buđenje, njene profesionalne interese, obrasce spavanja, biometrijske pokazatelje, pa čak i implicitne emocionalne reakcije. Mentor prati Lukinu krivulju učenja i preferirane stilove vizualizacije. Ova hiper-personalizacija temelji se na sposobnosti agenata da kontinuirano prikupljaju, integriraju i analiziraju ogromne količine multimodalnih podataka iz različitih izvora tijekom dugog vremenskog perioda. To zahtijeva napredne modele za korisničko modeliranje (User Modeling), razvoj dugoročne memorije za agente (koja nadilazi ograničenja standardnog kontekstnog prozora LLM-ova) i sofisticirane algoritme za kontinuirano učenje (Continual Learning) koji omogućuju agentima da se prilagođavaju promjenama u korisničkom ponašanju i preferencijama bez potrebe za potpunim retrainingom (Shamekhi et al., 2023; Chen & Liu, 2018). Naravno, ova razina personalizacije neizbjegivo povlači sa sobom kritična pitanja o privatnosti podataka, sigurnosti i potencijalu za manipulaciju, što zahtijeva robusne tehničke i regulatorne mehanizme zaštite (Susser et al., 2019).

### **8.3.4 Sinergija Rojeva**

Kolaborativni Multi-Agentski Sustavi: Vizija ne uključuje samo jednog sveznajućeg agenta po korisniku, već ekosustav specijaliziranih agenata koji međusobno surađuju. Tempo koordinira s Mentorom, Domusom, prometnim modulima, agentima kolega i prijatelja. Synapse, Scholar i Scribe surađuju na Aninom radnom zadatku. Ovaj koncept multi-agentskih sustava (MAS), ili "digitalnih kolektiva", omogućuje razlaganje složenih problema na manje, upravljive dijelove koje rješavaju specijalizirani agenti (Wooldridge, 2009; Hong et al., 2023). Ostvarenje ove vizije zahtijeva razvoj standardiziranih protokola za komunikaciju i interoperabilnost između agenata (čak i ako su razvijeni od strane različitih tvrtki), sofisticirane mehanizme za koordinaciju, pregovaranje i rješavanje konflikata unutar roja, te infrastrukturu sposobnu podržati masivnu paralelnu komunikaciju i računalne zahtjeve takvih distribuiranih sustava. Okviri poput AutoGen (Wu et al., 2023) i LangGraph (LangChain, 2024) predstavljaju rane korake u smjeru omogućavanja ovakve kompleksne agentne suradnje.

### **8.3.5 Čovjek-AI Simbioza: Ne samo Alat, već Partner**

Konačno, narativ sugerira pomak u samom odnosu čovjeka i umjetne inteligencije. Agenti nisu samo pasivni alati koje koristimo; oni postaju aktivni suradnici, mentori i posrednici. Anin radni proces nije zamjena njenog rada AI-jem, već simbioza gdje ona pruža strateško usmjerjenje i kritičku prosudbu, dok AI obavlja analitičke i generativne zadatke. Lukino učenje nije zamjena nastavnika, već dopuna kroz personaliziranog AI tutora. Društvene interakcije nisu zamijenjene, već posredovane i olakšane. Ovo ukazuje na budućnost gdje će ključna ljudska vještina biti ne samo korištenje AI alata, već sposobnost efikasne suradnje s AI partnerima, razumijevanje njihovih snaga i slabosti, te usmjeravanje njihovih sposobnosti prema željenim ciljevima (Dell'Acqua et al., 2023; Wilson & Daugherty, 2018).

Naravno, pokretačka snaga iza svih ovih trendova ostaje ona temeljna: kontinuirani napredak u računalnoj snazi (Poglavlje 5), razvoj sve većih i sposobnijih temeljnih modela (LLM-ova i šire) (Poglavlje 3), inovacije u algoritmima za učenje, rezoniranje i planiranje, te dostupnost ogromnih količina podataka potrebnih za treniranje i personalizaciju ovih sustava.

Konvergencija ovih pokretača čini viziju svijeta duboko isprepletene s AI agentima ne samo mogućom, već i sve vjerojatnjom u nadolazećem desetljeću.

## **8.4 GLASOVI BUDUĆNOSTI: PERSPEKTIVE MISLILACA I ISTRAŽIVAČA**

Predočena vizija svakodnevice u doba sveprisutnih AI agenata, sa svojom besprijeckornom učinkovitošću, dubokom personalizacijom i proaktivnom automatizacijom, ne predstavlja monolitnu ili univerzalnu prihvaćenu sliku budućnosti. Ona egzistira unutar širokog spektra predviđanja, nade i upozorenja koja dolaze od vodećih tehnologa, filozofa, društvenih znanstvenika i samih istraživača umjetne inteligencije. Analiza ovih različitih glasova ključna je za kritičko sagledavanje potencijalnih putanja i implikacija razvoja opisanih tehnologija.

S jedne strane, narativ o Aninom i Lukinom danu rezonira s optimističnim vizijama eksponencijalnog tehnološkog napretka, kakve artikulira Ray Kurzweil. Iz njegove perspektive, opisani stupanj integracije AI-ja u svakodnevni život, sposobnost agenata da anticipiraju potrebe i upravljaju kompleksnim sustavima, te simbioza čovjeka i stroja u radu i učenju predstavljaju logične korake na putanji prema tehnološkoj singularnosti – hipotetskoj točki u budućnosti kada će umjetna inteligencija nadmašiti ljudsku u svim aspektima, vodeći do nesagleđivih promjena u civilizaciji (Kurzweil, 2005). Trendovi poput ambijentalne inteligencije i personalizacije vide se kao preteće dubljeg spajanja biološke i nebiološke inteligencije, potencijalno vodeći prema transhumanističkim ishodima i radikalnom produljenju ljudskih sposobnosti. U ovoj viziji, agenti nisu samo pomoćnici, već partneri u transcendenciji bioloških ograničenja.

Nasuprot tome, ista narativna skica može se interpretirati kroz znatno oprezniju, pa čak i distopisku leću mislilaca poput Yuvala Noah Hararija ili Shoshane Zuboff. Harari upozorava na opasnost da algoritmi, hraneći se neprestanim protokom naših podataka, postanu sposobni "hakirati" ljudska bića – razumjeti nas bolje nego što mi razumijemo sami sebe – te suptilno manipulirati našim odlukama, željama i uvjerenjima (Harari, 2018). Sveprisutnost agenata koji optimiziraju naše rasporede, kuriraju naše informacije i čak analiziraju naša raspoloženja, kao u primjeru s Anom i Tempom, može se vidjeti kao ostvarenje scenarija u kojem ljudska agensnost i autonomija bivaju erodirane pod krinkom pogodnosti i efikasnosti. Slično tome, Zuboff (2019) bi u ovakvoj budućnosti prepoznala vrhunac "nadzornog kapitalizma", gdje se svaki aspekt ljudskog iskustva pretvara u bihevioralne podatke koji se koriste za predviđanje i modificiranje ponašanja u službi komercijalnih ili političkih ciljeva. Pitanje koje postavljaju jest: u svijetu gdje agenti upravljaju našim životima, tko upravlja agentima i s kojim ciljevima? Postajemo li korisnici ili proizvodi u algoritamski upravljanom društvu?

Gledajući prema još daljoj budućnosti i potencijalu umjetne opće inteligencije (AGI), perspektiva Nicka Bostroma unosi dodatni sloj egzistencijalne zabrinutosti. Iako Anini i Lukini agenti možda ne predstavljaju punu AGI, putanja razvoja koja omogućuje njihovu sofisticiranost i autonomiju neizbjježno vodi prema pitanjima o kontroli nad sustavima koji bi jednog dana mogli postati znatno inteligentniji od ljudi. Bostrom (2014) naglašava fundamentalni problem poravnjanja (alignment problem): kako osigurati da ciljevi i vrijednosti superinteligentnog AI sustava ostanu uskladjeni s ljudskim blagostanjem? Rizik da AGI razvije instrumentalne ciljeve koji su u sukobu s našim opstankom predstavlja, prema Bostromu, najveću egzistencijalnu prijetnju čovječanstvu. Čak i suradnja unutar "rojeva" agenata, prikazana u narativu, može postati nepredvidljiva i teška za kontrolu ako agenti dosegnu visoku razinu autonomije i sposobnosti samostalnog učenja i prilagodbe.

S pragmatičnije etičke i filozofske točke gledišta, Luciano Floridi skreće pozornost na neposrednije izazove koje postavljaju već postojeće i nadolazeće AI tehnologije, poput agenata opisanih u viziji. On naglašava potrebu za razvojem "etike za infosferu" – razumijevanja moralnih

pitanja koja proizlaze iz našeg života u svijetu prožetom informacijskim tehnologijama (Floridi, 2013; 2018). Problemi poput algoritamske pristranosti, transparentnosti, odgovornosti, zaštite privatnosti i digitalnog jaza nisu samo tehnička pitanja, već fundamentalni etički izazovi koji zahtijevaju promišljene governance okvire. Floridi bi vjerojatno analizirao agente poput Tempa ili Mentora ne nužno u terminima svijesti ili superinteligencije, već kao moćne izvore "informacijske frikcije" ili, obrnuto, manipulacije, koji preoblikuju naše epistemičke prakse (kako dolazimo do znanja), našu percepciju stvarnosti i naše međusobne odnose.

Konačno, važno je uključiti i glasove samih istraživača i inženjera koji aktivno rade na razvoju ovih tehnologija u vodećim akademskim i industrijskim laboratorijima (npr. OpenAI, Google DeepMind, Meta AI, Anthropic). Dok postoji značajan entuzijazam oko brzog napretka u sposobnostima modela (rezoniranje, multimodalnost, generiranje koda), unutar zajednice raste i svijest o ograničenjima i rizicima. Intenzivno se radi na problemima pouzdanosti, robusnosti, sigurnosti i smanjenju halucinacija (OpenAI, 2023; Anthropic, 2024). Pitanje svijesti u AI sustavima većina istraživača smatra otvorenim, ali generalno se slažu da današnji modeli, unatoč impresivnim performansama, ne posjeduju subjektivno iskustvo ili svijest u ljudskom smislu. Velik fokus stavlja se na tehničke aspekte poravnjanja – razvoj metoda poput učenja s pojačanjem iz ljudskih povratnih informacija (RLHF) ili novijih pristupa poput Constitutional AI (Bai et al., 2022) – kako bi se osiguralo da se modeli ponašaju na način koji je koristan i bezopasan. Iako se vode rasprave o tome koliko smo blizu AGI-ju, postoji sve veći konsenzus o potrebi za odgovornim razvojem i proaktivnim razmatranjem potencijalnih društvenih posljedica (Bengio et al., 2023).

Ovaj spektar perspektiva – od tehnološkog optimizma, preko humanističke zabrinutosti i egzistencijalnog opreza, do pragmatičnih etičkih razmatranja i tehničkih izazova – ostikava kompleksnost i neizvjesnost budućnosti s naprednom umjetnom inteligencijom. Vizija digitalnih suputnika koji orkestriraju našu svakodnevnicu nije ni neizbjegna utopija ni zajamčena distopija; ona je potencijalna budućnost čiji će konačni oblik ovisiti o tehnološkim probojima, ali još više o društvenim izborima, etičkim promišljanjima i regulatornim odlukama koje donosimo danas.

## 8.5 DRUŠTVO U TRANSFORMACIJI: NAZNAKE NOVIH NORMA I IZAZOVA

Vizija svijeta prožetog sveprisutnim i autonomnim AI agentima, kako je skicirana u prethodnim odjeljcima, implicira transformacije koje daleko nadilaze individualna iskustva pogodnosti ili nelagode. Takva duboka integracija umjetne inteligencije u svakodnevni život neizbjegno bi dovela do fundamentalnih promjena u samoj strukturi društva, preoblikujući uspostavljene norme, institucije i međuljudske odnose. Analiza potencijalnih društvenih posljedica ključna je za anticipiranje izazova i usmjeravanje razvoja prema ishodima koji su usklađeni s ljudskim vrijednostima. Nekoliko ključnih područja transformacije nameće se kao posebno značajno.

Prvo, svijet rada doživio bi seizmičku promjenu. Automatizacija kognitivnih zadataka, koju agenci poput Synapsea, Scholara i Scribea demonstriraju u Aninom radnom okruženju, proširila bi se na brojne profesije, potencijalno dovodeći do značajne disruptcije na tržištu rada. Dok optimistični pogledi naglašavaju augmentaciju ljudskih sposobnosti i stvaranje novih radnih mјesta usmjerenih na suradnju s AI (npr. AI supervizori, etički revizori, prompt inženjeri visoke razine) (Wilson & Daugherty, 2018; Dell'Acqua et al., 2023), zabrinutost oko masovne nezaposlenosti u određenim sektorima i potrebe za radikalnom prekvalifikacijom radne snage ostaje opravdana (Frey & Osborne, 2017; Acemoglu & Restrepo, 2020). Sama priroda traženih vještina vjerojatno bi se promijenila, s manjim naglaskom na rutinsko procesiranje informacija,

a većim na kritičko razmišljanje, kreativnost, emocionalnu inteligenciju i sposobnost upravljanja kompleksnim čovjek-Al sustavima. Ove promjene mogle bi potaknuti i šire debate o ekonomskim modelima, uključujući rasprave o univerzalnom temeljnem dohotku (UBI) kao mogućem odgovoru na tehnološku nezaposlenost.

Drugo, sustav obrazovanja prošao bi kroz korjenitu transformaciju. Personalizirano učenje, kakvo doživljava Luka uz pomoć 'Mentora', moglo bi postati norma, a ne iznimka. To obećava ispunjenje dugogodišnjeg pedagoškog idealja o prilagodbi obrazovanja individualnim potrebama svakog učenika (VanLehn, 2011). Međutim, to također implicira fundamentalnu promjenu uloge ljudskih nastavnika. Oni bi se sve više pomicali od primarnih prenositelja informacija prema ulogama mentora, facilitatora, dizajnera obrazovnih iskustava i stručnjaka za razvoj socijalno-emocionalnih vještina – aspekata koje AI teško može replicirati (Zawacki-Richter et al., 2019). Kurikulumi bi se morali prilagoditi, stavlјajući veći naglasak na digitalnu pismenost, kritičko vrednovanje informacija (posebno onih generiranih od strane AI), etiku umjetne inteligencije i vještine suradnje s AI sustavima. Izazov bi bio osigurati da ova AI-poboljšana edukacija bude dostupna svima, kako se ne bi produbio postojeći obrazovni jaz.

Treće, sama priroda međuljudskih odnosa i društvene povezanosti mogla bi se redefinirati. Posredovanje agenata u komunikaciji, kao što Tempo pomaže Ani u organizaciji druženja i filtriranju društvenih informacija, nudi efikasnost, ali postavlja pitanja o autentičnosti i kvaliteti veza. Postoji rizik da preveliko oslanjanje na agente za upravljanje društvenim životom doveđe do atrofije socijalnih vještina ili stvaranja površnjih odnosa (Turkle, 2017). Mogućnost da agenti analiziraju i interpretiraju emocionalno stanje korisnika i njihovih kontakata, iako potencijalno korisna, otvara vrata suptilnoj manipulaciji i eroziji emocionalne privatnosti. Nadalje, mogli bismo svjedočiti pojavi sve kompleksnijih parasocijalnih, pa čak i kvazi-emocionalnih veza između ljudi i njihovih AI agenata, što postavlja filozofska pitanja o prirodi odnosa i empatije u doba sintetičkih sugovornika (Przybylski & Weinstein, 2013; Baecker & Kospoth, 2022).

Četvrto, koncept privatnosti suočio bi se s dosad neviđenim izazovima. Hiper-personalizacija i ambijentalna inteligencija, koje pokreću agente poput Tempa i Domusa, inherentno zahtijevaju kontinuirano prikupljanje i analizu ogromnih količina osobnih, bhevioralnih i biometrijskih podataka. U svijetu gdje su senzori sveprisutni, a agenti stalno "slušaju" i "promatraju" kako bi anticipirali naše potrebe, tradicionalni mehanizmi pristanka i kontrole postaju nedostatni (Zuboff, 2019; Susser et al., 2019). Osiguravanje smislenog korisničkog nadzora nad podacima, implementacija robusnih tehnika za očuvanje privatnosti (poput diferencijalne privatnosti ili federiranog učenja tamo gdje je primjenjivo) i uspostavljanje snažnih regulatornih okvira postat će apsolutni imperativi kako bi se sprječilo stvaranje sveobuhvatnog nadzornog aparata, bilo korporativnog ili državnog.

Peto, postoji značajan rizik da će ove tehnologije, ako se ne upravljaju pažljivo, produbiti postojeće društvene i ekonomске nejednakosti. Digitalni jaz mogao bi se proširiti, stvarajući podjelu ne samo na one koji imaju pristup internetu, već i na one koji si mogu priuštiti najnaprednije, personalizirane AI agente i one koji ostaju s osnovnim ili nikakvim verzijama. Nadalje, ako se ne ulože svjesni napor u detekciju i mitigaciju algoritamske pristranosti, agenti bi mogli perpetuirati ili čak pojačati postojeće stereotipe i diskriminaciju u područjima poput zapošljavanja, kreditiranja ili čak obrazovnih preporuka (Bender et al., 2021; Noble, 2018). Istovremeno, postoji i kontrapotencijal: AI agenti bi mogli djelovati kao sila demokratizacije, pružajući pristup kvalitetnim informacijama, personaliziranom obrazovanju ili osnovnim zdravstvenim savjetima populacijama koje su tradicionalno bile uskraćene za takve resurse.

Ključno će biti osigurati pravičan pristup i dizajnirati sustave koji aktivno rade na smanjenju, a ne povećanju, nejednakosti.

Ove naznake transformacije – u radu, obrazovanju, odnosima, privatnosti i jednakosti – nisu izolirani fenomeni. One su međusobno povezane i vjerojatno će stvarati kompleksne povratne sprege. Društvo budućnosti, oblikovano sveprisutnim AI agentima, vjerojatno će zahtijevati nove društvene ugovore, redefinirane etičke norme i adaptivne oblike upravljanja kako bi se uspješno nosilo s izazovima i iskoristilo prilike koje ova moćna tehnologija donosi.

## 8.6 ZAKLJUČAK POGLAVLJA: IZMEĐUUTOPIJE I DISTOPIJE – OTVORENA PITANJA

## 9 REFERENCE

- Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Ahmed, A., Alaboudi, A. A., Zizi, M., ... & Househ, M. (2022). Perceptions and opinions of patients about mental health chatbots: Scoping review. *Journal of Medical Internet Research*, 24(1), e32 perceptions. [https://doi.org/10.2196/32\\_perceptions](https://doi.org/10.2196/32_perceptions)
- Aboujaoude, E., Savage, M. W., Starcevic, V., & Salame, W. O. (2015). Cyberbullying: Review of an old problem gone viral. *Journal of Adolescent Health*, 57(1), 10-18.
- Acampora, G., Cook, D. J., Rashidi, P., & Vasilakos, A. V. (2013). A survey on ambient intelligence in health care. *Journal of Ambient Intelligence and Humanized Computing*, 4(4), 365-394. <https://doi.org/10.1109/JPROC.2013.2262913>
- Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6), 2188-2244.
- Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6), 2188-2244.
- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.
- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.
- Adam, M., Wessel, M., Benlian, A., & Hess, T. (2021). *AI-based chatbots in customer service and their effects on user compliance*. *Electronic Markets*, 31(2), 427–445.
- Agrawal, R., Kumar, A., Subbian, K., & Adluru, V. (2023). Large Language Models for E-commerce: A Survey Framework. *arXiv preprint arXiv:2309.08231*.
- Aiello, L. C., & Wheeler, P. (1995). *The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution*. *Current Anthropology*, 36(2), 199–221.
- Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction Based on Adaptive Learning Technologies. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 522-560). Routledge.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236.

- Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in the 2016 Election*. *Journal of Economic Perspectives*, 31(2), 211–236.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-36.
- Alter, A. (2017). Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked. Penguin Press.
- Altmann, J., & Sauer, F. (2017). *Autonomous Weapon Systems and Strategic Stability*. Survival, 59(5), 117–142.
- AMD. (2023, December 6). *AMD Instinct™ MI300 Series Accelerators*. AMD. [Potražiti najnoviju specifikaciju ili press release za MI300X/A]
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Andrejevic, M. (2007). Surveillance in the digital enclosure. *The Communication Review*, 10(4), 295–317.
- Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), 185–205.
- Anthropic. (2023). Claude 3. Preuzeto s: [anthropic.com](https://anthropic.com)
- Anthropic. (2023). Claude 3. Preuzeto s: <https://www.anthropic.com/>
- Anthropic. (2023). Claude. Preuzeto s: <https://www.anthropic.com>
- Anthropic. (2023). Claude. Preuzeto s: [https://www.anthropic.com/](https://www.anthropic.com)
- Anthropic. (2024, March 4). Introducing the next generation of Claude. Anthropic News. <https://www.anthropic.com/news/clause-3-family>
- Anthropic. (2024, March 4). *Introducing the next generation of Claude*. Anthropic News. <https://www.anthropic.com/news/clause-3-family>
- Anthropic. (2024, March 4). *Introducing the next generation of Claude*. Anthropic News. <https://www.anthropic.com/news/clause-3-family> (Često uključuju i izjave o sigurnosti i etici).
- Apple, M. W. (2004). *Ideology and Curriculum*. RoutledgeFalmer.
- Apple, M. W. (2004). *Ideology and Curriculum*. RoutledgeFalmer.
- Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., ... & Martinis, J. M. (2019). *Quantum supremacy using a programmable superconducting processor*. *Nature*, 574(7779), 505–510.

- AWS. (2023, November 28). *Announcing the next generation of AWS-designed chips*. AWS News Blog. <https://aws.amazon.com/blogs/aws/announcing-the-next-generation-of-aws-designed-chips/>
- AWS. (2024). *Amazon EC2 UltraClusters for distributed training*. [Potražiti najnoviju AWS dokumentaciju]
- Azevedo, R., & Aleven, V. (Eds.). (2013). International handbook of metacognition and learning technologies. Springer.
- Baecker, T., & Kospoth, T. (2022). Parasocial Relations with AI Companions: Concepts, Predictors, and Consequences of User–AI Companion Relationship Formation. *Human-Computer Interaction*, 1–29. <https://doi.org/10.1080/07370024.2022.2138878>
- Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., ... & Amodei, D. (2022). *Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback*. arXiv preprint arXiv:2204.05862.
- Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., ... & Amodei, D. (2022). *Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback*. arXiv preprint arXiv:2204.05862.
- Bai, Y., Jones, A., Ndousse, K., et al. (2022). Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. *arXiv:2204.05862*.
- Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI Feedback*. arXiv preprint arXiv:2212.08073. <https://arxiv.org/abs/2212.08073>
- Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI Feedback*. arXiv preprint arXiv:2212.08073.
- Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
- Barnett, S., et al. (2024). *Seven Failure Points When Engineering a Retrieval Augmented Generation System*. arXiv preprint arXiv:2401.05856. (Pruža uvid u praktične izazove RAG-a).
- Baron, N. S. (2008). *Always On: Language in an Online and Mobile World*. Oxford University Press.
- Bartky, I. R. (2007). *One Time Fits All: The Campaigns for Global Uniformity*. Stanford University Press.
- Bartky, I. R. (2007). *One Time Fits All: The Campaigns for Global Uniformity*. Stanford University Press.
- Bartky, I. R. (2007). *One Time Fits All: The Campaigns for Global Uniformity*. Stanford University Press.
- Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2019). A review on deep learning for recommender systems: challenges and remedies. *Artificial Intelligence Review*, 52(1), 1-37. <https://doi.org/10.1007/s10462-018-9654-y>

- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191.
- Bawden, D., & Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180-191.
- Bell, A. G. (1876). Improvement in telegraphy. U.S. Patent 174465.
- Bell, A. G. (1876). *Improvement in Telegraphy. U.S. Patent 174465*.
- Bell, A. G. (1876). *Improvement in Telegraphy*. U.S. Patent No. 174,465.
- Bendel, O. (2019). From Chatbot to Social Bot: Conversational Agents and Their Perception by Users. In *Digital Transformation in Journalism and News Media* (pp. 147–162). Springer.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *ACL 2020*.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *ACL 2020*.
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *ACL 2020*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?.* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FACCT 2021*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FACCT 2021*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FACCT '21*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? **FACCT '21**.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?  . Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
- Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Maharaj, T., ... & Zhang, R. (2023). *Managing AI Risks in an Era of Rapid Progress*. arXiv preprint arXiv:2310.17688. <https://arxiv.org/abs/2310.17688> (Izjava grupe vodećih istraživača).
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
- Bennett, W. L., & Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press.
- Bennett, W. L., & Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge University Press.
- Bernard, J. (1993). *The Idea of Monasticism: A Critical Edition*. Princeton University Press.
- Bernard, R. (1993). *The Monastic Scribal Tradition*. Oxford University Press.
- Bernard, R. (1993). *The Monastic Scribal Tradition*. Oxford University Press.
- Berners-Lee, T. (1990). *Information Management: A Proposal*. CERN.
- Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). *Quantum machine learning*. *Nature*, 549(7671), 195–202.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of “Bias” in NLP. *ACL 2020*.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of “Bias” in NLP. *ACL 2020*.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of “Bias” in NLP. *ACL 2020*.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of “Bias” in NLP. *ACL 2020*.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of “Bias” in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 54–79.
- Blondheim, M. (1994). *News Over the Wires: The Telegraph and the Flow of Public Information in America, 1844–1897*. Harvard University Press.
- Blondheim, M. (1994). *News Over the Wires: The Telegraph and the Flow of Public Information in America, 1844–1897*. Harvard University Press.
- Blondheim, M. (1994). *News Over the Wires: The Telegraph and the Flow of Public Information in America, 1844–1897*. Harvard University Press.

- Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.
- Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zhang, X. (2016). *End to End Learning for Self-Driving Cars*. arXiv preprint arXiv:1604.07316.
- Boltz, W. (1994). *The Origin and Early Development of the Chinese Writing System*. American Oriental Society.
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. Advances in Neural Information Processing Systems, 4349–4357.
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. Advances in Neural Information Processing Systems, 29, 4349–4357.
- Bommasani, R. et al. (2021). On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*.
- Bommasani, R. et al. (2021). On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). *On the Opportunities and Risks of Foundation Models*. arXiv preprint *arXiv:2108.07258*.
- Bommasani, R., Hudson, D., Adeli, E., et al. (2021). On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258*.
- Bonk, C. J. (2009). *The World Is Open: How Web Technology Is Revolutionizing Education*. Jossey-Bass.
- Bonk, C. J. (2009). *The World is Open: How Web Technology is Revolutionizing Education*. Jossey-Bass.
- Boroditsky, L. (2011). *How Language Shapes Thought*. Scientific American, 304(2), 62–65.
- Boroditsky, L. (2011). *How Language Shapes Thought*. Scientific American, 304(2), 62–65.
- Boroditsky, L. (2011). How Languages Construct Time. *Annual Review of Psychology*, 62, 61–86.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bourdieu, P. (1990). *The Logic of Practice*. Stanford University Press.
- Bourdieu, P. (1990). *The Logic of Practice*. Stanford University Press.

- Boyd, D. (2006). Friends, friendsters, and MySpace top 8: Writing community into being on social network sites. *First Monday*, 11(12).
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Boyd, d. m., & Ellison, N. B. (2007). *Social Network Sites: Definition, History, and Scholarship*. *Journal of Computer-Mediated Communication*, 13(1), 210–230.
- Brandtzæg, P. B., Folstad, A., & Lutnæs, E. (2022). Social Chatbots: Emotions and Trust in Conversational User Interfaces. *International Journal of Human-Computer Interaction*.
- Brants, K., & Voltmer, K. (Eds.). (2011). *Political Communication in Postmodern Democracy: Challenging the Primacy of Politics*. Palgrave Macmillan.
- Briggs, A., & Burke, P. (2009). *A Social History of the Media: From Gutenberg to the Internet*. Polity Press.
- Briggs, A., & Burke, P. (2009). *A Social History of the Media: From Gutenberg to the Internet*. Polity Press.
- Brown, T. B., Mann, B., Ryder, N., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language Models are Few-Shot Learners. *NeurIPS 2020*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems*, 33, 1877–1901.

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners*. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners*. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners*. *Advances in Neural Information Processing Systems*, 33, 1877–1901. (NeurIPS 2020)
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners*. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language Models are Few-Shot Learners*. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. arXiv preprint arXiv:1802.07228.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at Work. NBER Working Paper No. 31161. <https://www.nber.org/papers/w31161>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI and the workforce. *Science*, 381(6659), 710–712. <https://doi.org/10.1126/science.adl9360>
- Buchanan, B., Lohn, A. J., Musser, M., & Sedova, K. (2021). Truth, Lies, and Automation: How Language Models Could Change Disinformation. Center for Security and Emerging Technology. <https://cset.georgetown.edu/publication/truth-lies-and-automation/>
- Buchanan, B., Lonergan, K., Kwik, G., Kempton, J., & Caplan, A. (2021). *Truth, Lies, and Automation: How Language Models Could Change Disinformation*. Center for Security and Emerging Technology.
- Buckner, R. L., & Krienen, F. M. (2013). *The Evolution of Distributed Association Networks in the Human Brain*. *Trends in Cognitive Sciences*, 17(12), 648–665.
- Bullmore, E., & Sporns, O. (2012). *The Economy of Brain Network Organization*. *Nature Reviews Neuroscience*, 13(5), 336–349.
- Burke, P. (2000). *A Social History of Knowledge: From Gutenberg to Diderot*. Polity Press.
- Burns, R. W. (2000). *Television: An International History of the Formative Years*. IET.
- Burns, R. W. (2000). *Television: An International History of the Formative Years*. The Institution of Engineering and Technology.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.

- Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*.
- Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*.
- Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*.
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Calvo, R. A., D'Mello, S., Gratch, J., & Kappas, A. (2018). *The Oxford Handbook of Affective Computing*. Oxford University Press.
- Cambria, E., & White, B. (2014). *Jumping NLP Curves: A Review of Natural Language Processing Research*. IEEE Computational Intelligence Magazine, 9(2), 48–57.
- Campbell, J. (1949). *The Hero with a Thousand Faces*. Princeton University Press.
- Campbell, J. (1949). *The Hero with a Thousand Faces*. Princeton University Press.
- Campbell, J. (1949). *The Hero with a Thousand Faces*. Princeton University Press.
- Campbell, J. (1949). *The Hero with a Thousand Faces*. Princeton University Press.
- Cantril, H. (1940). *The Invasion from Mars: A Study in the Psychology of Panic*. Princeton University Press.
- Carey, J. W. (1989). *Communication as Culture: Essays on Media and Society*. Routledge.
- Carey, J. W. (1989). *Communication as Culture: Essays on Media and Society*. Routledge.
- Carey, J. W. (1989). *Communication as Culture: Essays on Media and Society*. Routledge.
- Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting Training Data from Large Language Models. *Proceedings of the 30th USENIX Security Symposium*, 2633-2650.
- Casal, G., Costa-jussà, M. R., & Fonollosa, J. A. R. (2023). Exploring Emotional Adaptation in Conversational AI: A Multimodal Approach. *EMNLP 2023 (Accepted paper)*.

- Casper, S., Davies, X., Shi, C., Gilbert, T. K., Scheurer, J., Rando, J., ... & Hadfield-Menell, D. (2023). *Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback*. arXiv preprint arXiv:2307.15217.
- Castells, M. (1996). *The Rise of the Network Society*. Blackwell Publishers.
- Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press.
- Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press.
- Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press.
- Castells, M. (2015). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Polity Press.
- Castells, M. (2015). *Networks of Outrage and Hope: Social Movements in the Internet Age*. Polity Press.
- Castells, M. (2015). *Networks of Outrage and Hope: Social Movements in the Internet Age* (2nd ed.). Polity Press.
- Chandler, A. D., & Cortada, J. W. (2000). *A Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*. Oxford University Press.
- Chandler, A. D., & Cortada, J. W. (2000). *A Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*. Oxford University Press.
- Chandler, A. D., & Cortada, J. W. (2000). *A Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*. Oxford University Press.
- Chandler, D. (2007). *Semiotics: The Basics*. Routledge.
- Chandler, D. (2007). *Semiotics: The Basics*. Routledge.
- Chaudhuri, S., & Boonthum-Denecke, C. (2018). Conversational Agents for Language Learning: A Systematic Review. *CALICO Journal*, 35(1), 1–29.
- Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. *Proceedings of the IEEE International Conference on Computer Vision*, 2722–2730.
- Chen, Z., & Liu, B. (2018). *Lifelong Machine Learning* (2nd ed.). Morgan & Claypool Publishers.
- Cheng, W., Wang, M., Chau, C., & Hui, P. (2023). Is ChatGPT the Ultimate Teammate? How AI-Assisted Collaboration Impacts Creativity. arXiv preprint arXiv:2305.08319.
- Chi, H., Zhang, X., Qin, Y., Yu, H., & Mei, Q. (2021). Improving Educational Dialogue Generation via Context-Specific Knowledge Extraction. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 729–740.

- Chi, H., Zhang, X., Qin, Y., Yu, H., & Mei, Q. (2021). *Improving Educational Dialogue Generation via Context-Specific Knowledge Extraction*. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 729–740.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Dean, J. (2022). *PaLM: Scaling Language Modeling with Pathways*. arXiv preprint arXiv:2204.02311.
- Christiano, P. F., Leike, J., Brown, T., et al. (2017). Deep Reinforcement Learning from Human Preferences. *NeurIPS 2017*.
- Ciocca, G., Napoletano, P., & Schettini, R. (2023). Recent advances in deep learning for personalized recommendation systems: A survey. *Journal of King Saud University-Computer and Information Sciences*, 35(8), 101641. <https://doi.org/10.1016/j.jksuci.2023.101641>
- Clanchy, M. T. (1993). *From Memory to Written Record: England 1066–1307*. Blackwell.
- Coe, L. (1993). *The Telegraph: A History of Morse's Invention and Its Predecessors in the United States*. McFarland & Company.
- Coe, L. (1993). *The Telephone and Its Several Inventors: A History*. McFarland.
- Coe, L. (1993). *The Telephone and Its Several Inventors: A History*. McFarland.
- Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.
- Collins, F. S., & Varmus, H. (2015). *A New Initiative on Precision Medicine*. New England Journal of Medicine, 372(9), 793-795.
- Collins, F. S., & Varmus, H. (2015). *A New Initiative on Precision Medicine*. New England Journal of Medicine, 372(9), 793-795.
- Companion, M. H., Lee, M., & Agrawala, M. (2023). CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Integration in the Creative Process. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Article 418, 1–17. <https://doi.org/10.1145/3544548.3581228>
- Companion, M., Kampman, O., Zhu, B., Ramesh, A., & Child, R. (2023). Suno and Udio: Exploring AI Music Generation Systems. [Potražiti recenzije ili tehničke opise ovih sustava iz 2024.]
- Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., He, H., Hwang, W., ... & Zada, S. (2022). No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv preprint arXiv:2207.04672.
- Cotterrell, R. (2017). *Law and Language: A Critical Introduction*. Routledge.
- Cotterrell, R. (2017). *Law and Language: A Critical Introduction*. Routledge.
- Coulmas, F. (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge University Press.
- Cowie, R. (2015). *Ethical Issues in Affective Computing*. U *The Oxford Handbook of Affective Computing* (str. 334–348). Oxford University Press.

- Cozzi, A., Marotta, R., Paone, M., Patruno, A., & Spinelli, M. (2020). *Massively Parallel Computing*. In: High-Performance Computing on Complex Environments. Wiley.
- Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Routledge.
- Cui, G., & Zhang, X. (2024). Large Language Models for Business Intelligence: A Survey. arXiv preprint arXiv:2402.12673. <https://arxiv.org/abs/2402.12673>
- Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: Transparency and collaboration in an open software repository. *CSCW '12*.
- Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: transparency and collaboration in an open software repository. *Proceedings of the ACM 2012 conference on computer supported cooperative work*, 1277-1286.
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811-817. (Daje pregled povijesti i povratka chatbotova).
- Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., Cao, Y., Choday, S. H., ... & Wang, H. (2018). *Loihi: A Neuromorphic Manycore Processor with On-Chip Learning*. *IEEE Micro*, 38(1), 82–99.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Penguin.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.
- Dehaene, S., Lau, H., & Kouider, S. (2017). *What is consciousness, and could machines have it?* *Science*, 358(6362), 486–492.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013*.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013*. <https://ssrn.com/abstract=4573321>
- Deng, X., Wang, G., & Sun, Y. (2019). *Financial fraud detection using heterogeneous information*. *IEEE Transactions on Knowledge and Data Engineering*, 32(9), 1822–1835.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 4171–4186.

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 4171–4186.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv:1810.04805.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 4171–4186.
- Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Company.
- Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton.
- Dixon, M., Halperin, I., & Bilokon, P. (2020). *Machine Learning in Finance: From Theory to Practice*. Springer.
- Dowling, M., & Zaki, M. (2023). ChatGPT for scientific writing: the disruptor we need?. *Journal of Clinical Oncology*, 41(18), 3373.
- Dugan, J., Peskov, D., Sargsyan, N., & Nenkova, A. (2022). Understanding and Reducing Toxicity Hallucinations in Language Models. *EMNLP 2022*.
- Duolingo. (2023, March 14). Introducing Duolingo Max, a learning experience powered by GPT-4. Duolingo Blog. <https://blog.duolingo.com/duolingo-max/>
- Duolingo. (2023, March 14). *Introducing Duolingo Max, a learning experience powered by GPT-4*. Duolingo Blog. <https://blog.duolingo.com/duolingo-max/>
- Durkheim, E. (1912). *Les Formes élémentaires de la vie religieuse*. Alcan.
- Durkheim, E. (1912). *The Elementary Forms of Religious Life*. Allen & Unwin.
- Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge University Press.
- Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge University Press.
- Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge University Press.
- Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge University Press.
- El Ayadi, M., Kamel, M. S., & Karray, F. (2011). *Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases*. *Pattern Recognition*, 44(3), 572–587.

- Elgammal, A. (2022). AI is blurring the definition of artist. *American Scientist*, 110(3), 184.
- Elgammal, A., Liu, B., Kim, D., Elhoseiny, M., & Mazzone, M. (2017). *CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. Proceedings of the 8th International Conference on Computational Creativity*, 96–103.
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends”: Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143–1168.
- Elman, J. L. (1990). *Finding Structure in Time. Cognitive Science*, 14(2), 179–211.
- Epstein, Z., Hertzmann, A., Akten, M., Farid, H., Fjeld, J., Frank, M. R., ... & Russakovsky, O. (2023). Art and the science of generative AI. *Science*, 380(6650), 1110-1111.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). *Dermatologist-level classification of skin cancer with deep neural networks. Nature*, 542(7639), 115–118.
- European Parliament. (2024, March 13). Artificial intelligence (AI) act: MEPs adopt landmark law. European Parliament News. <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-ai-act-meps-adopt-landmark-law>
- Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Routledge.
- Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Routledge.
- Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Routledge.
- Fan, A., Bhosale, S., Schwenk, H., Ma, M., El-Kishky, A., Goyal, N., ... & Edunov, S. (2021). *Beyond English-Centric Multilingual Machine Translation. Journal of Machine Learning Research*, 22(107), 1–48.
- Fan, A., Bhosale, S., Schwenk, H., Ma, M., El-Kishky, A., Goyal, N., ... & Edunov, S. (2021). *Beyond English-Centric Multilingual Machine Translation. Journal of Machine Learning Research*, 22(107), 1–48.
- Fan, A., Bhosale, S., Schwenk, H., Ma, M., El-Kishky, A., Goyal, N., ... & Edunov, S. (2021). *Beyond English-Centric Multilingual Machine Translation. Journal of Machine Learning Research*, 22(107), 1–48.
- Fan, A., Bhosale, S., Schwenk, H., Ma, M., El-Kishky, A., Goyal, N., ... & Edunov, S. (2021). *Beyond English-Centric Multilingual Machine Translation. Journal of Machine Learning Research*, 22(107), 1–48.
- Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., ... & Joulin, A. (2021). *Beyond English-Centric Multilingual Machine Translation. Journal of Machine Learning Research*, 22(107), 1–48.

- Febvre, L., & Martin, H.-J. (1976). *The Coming of the Book: The Impact of Printing 1450–1800*. Verso.
- Febvre, L., & Martin, H.-J. (1976). *The Coming of the Book: The Impact of Printing 1450–1800*. Verso.
- Filliozat, P. (2004). Ancient Sanskrit Mathematics: An Oral Tradition and a Written Literature. In C. J. Tuplin & T. E. Rihll (Eds.), *Science and Mathematics in Ancient Greek Culture* (pp. 138–159). Oxford University Press. (Raspravlja o usmenoj prirodi i metodama u sanskrtskim tradicijama).
- Finn, C., Abbeel, P., & Levine, S. (2017). *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*. Proceedings of the 34th International Conference on Machine Learning, 1126–1135.
- Finn, C., Abbeel, P., & Levine, S. (2017). *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*. Proceedings of the 34th International Conference on Machine Learning, 1126–1135.
- Finnegan, R. (1970). *Oral Literature in Africa*. Oxford University Press.
- Fischer, C. (1992). *America Calling: A Social History of the Telephone to 1940*. University of California Press.
- Fischer, C. (1992). *America Calling: A Social History of the Telephone to 1940*. University of California Press.
- Fischer, C. S. (1992). *America Calling: A Social History of the Telephone to 1940*. University of California Press.
- Fitzpatrick, K. K., Darcy, A. M., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot). *JMIR Mental Health*, 4(2).
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). *Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial*. *JMIR Mental Health*, 4(2), e19.
- Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180081. <https://doi.org/10.1098/rsta.2018.0081>
- Floridi, L., & Chiriatti, M. (2020). *GPT-3: Its Nature, Scope, Limits, and Consequences*. *Minds and Machines*, 30(4), 681–694.
- Floridi, L., & Chiriatti, M. (2020). *GPT-3: Its Nature, Scope, Limits, and Consequences*. *Minds and Machines*, 30(4), 681–694.
- Floridi, L., & Cowls, J. (2019). *A Unified Framework of Five Principles for AI in Society*. *Harvard Data Science Review*, 1(1).

- Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.
- Ford, M. (2018). *Architects of Intelligence: The Truth About AI from the People Building It*. Packt Publishing.
- Foucault, M. (1972). *The Archaeology of Knowledge*. Pantheon Books.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, 114, 254-280.
- Frey, C. B., & Osborne, M. A. (2017). *The Future of Employment: How Susceptible are Jobs to Computerisation? Technological Forecasting and Social Change*, 114, 254–280.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, 114, 254-280.
- Fung, P., Bertero, D., Liu, R., & Madotto, A. (2020). Empathetic and Emotional Dialogue Systems. *The Handbook of Multimodal-Multisensor Interfaces*, 3, ACM.
- Gao, L., Ma, X., Lin, J., & Callan, J. (2022). *Precise Zero-Shot Dense Retrieval without Relevance Labels*. arXiv preprint arXiv:2212.10496. (HyDE tehnika).
- Gao, S., Ge, S., Chen, F., & Sun, M. (2021). *Open Domain Conversation Generation with Latent Images*. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14), 12801–12809.
- Gao, S., Ge, S., Chen, F., & Sun, M. (2021). *Open Domain Conversation Generation with Latent Images*. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14), 12801–12809.
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Sun, H. (2024). Retrieval-Augmented Generation for Large Language Models: A Survey. *ACM Transactions on Intelligent Systems and Technology*, 15(2), 1-58. (Sveobuhvatan pregled RAG tehnika).
- Garcia, A., Xia, Y., Chang, S., & Finn, C. (2022). *Towards Cross-Cultural Understanding of Language Models*. arXiv preprint arXiv:2206.03364.
- Garcia, A., Xia, Y., Chang, S., & Finn, C. (2022). *Towards Cross-Cultural Understanding of Language Models*. arXiv preprint arXiv:2206.03364.
- Garcia-Ceja, E., Osmani, V., & Mayora, O. (2016). *Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step*. IEEE Journal of Biomedical and Health Informatics, 20(4), 1053–1060.
- Geng, S., Liu, J., Zhu, Y., Zhao, W. X., & Wen, J. R. (2022). Recommendation as Instruction Following: A Large Language Model-Empowered Recommendation Approach. *arXiv preprint arXiv:2212.01194*. <https://arxiv.org/abs/2212.01194>
- Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. Pluto Press.
- Gilbert, S., Rot R., & Berkowitz S. (2023). Charting the Course: Responsible Integration of Large Language Models in Health Care. *NEJM Catalyst Innovations in Care Delivery*, 4(6). <https://doi.org/10.1056/CAT.23.0246>

- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-193). MIT Press. (Raspravlja o algoritmima kao kuratorima).
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press. (Iako iz 2018., ostaje temeljni rad o moderaciji).
- GitHub. (2024). GitHub Copilot. GitHub Features. <https://github.com/features/copilot>
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Van Essen, D. C. (2016). A Multi-modal Parcellation of Human Cerebral Cortex. *Nature*, 536(7615), 171–178.
- Gleick, J. (2011). *The Information: A History, a Theory, a Flood*. Pantheon Books.
- Gleick, J. (2011). *The Information: A History, a Theory, a Flood*. Pantheon.
- Gleick, J. (2011). *The Information: A History, a Theory, a Flood*. Pantheon.
- Goertzel, B. (2014). *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*. Journal of Artificial General Intelligence, 5(1), 1–48.
- Goertzel, B., & Pennachin, C. (2007). *Artificial General Intelligence*. Springer.
- Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Pantheon Books.
- Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. Pantheon Books.
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Sedova, K., & Gentzel, M. (2023). Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. *arXiv preprint arXiv:2301.03773*.
- Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Sedova, K., & Lohn, A. J. (2023). Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. OpenAI & Stanford Internet Observatory. <https://cdn.openai.com/papers/generative-language-models-and-automated-influence-operations.pdf>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goody, J. (1987). *The Interface Between the Written and the Oral*. Cambridge University Press.
- Goody, J. (1987). *The Interface Between the Written and the Oral*. Cambridge University Press.
- Google Cloud. (2023, August 29). *Cloud TPU v5p and AI software advancements optimize for cost-efficient scale*. Google Cloud

Blog. <https://cloud.google.com/blog/products/compute/google-cloud-tpu-v5p-ai-supercomputer>

- Google DeepMind. (2023). Gemini: General-Purpose AI Model. Preuzeto s: <https://blog.google/technology/ai/>
- Google DeepMind. (2023). Gemini: General-Purpose AI Model. Preuzeto s: <https://blog.google/technology/ai/>
- Google DeepMind. (2023b, December 14). FunSearch: Making discoveries in mathematical sciences using Large Language Models. Google DeepMind Blog. <https://deepmind.google/discover/blog/funsearch-making-discoveries-in-mathematical-sciences-using-large-language-models/>
- Google. (2024). Gemini Models. Google DeepMind. [Potražiti najnoviju dokumentaciju ili najave]
- Greenhalgh, T. (2014). *How to Read a Paper: The Basics of Evidence-based Medicine*. BMJ Books.
- Greenhalgh, T. (2014). *How to Read a Paper: The Basics of Evidence-based Medicine*. BMJ Books.
- Greenhalgh, T. (2014). *How to Read a Paper: The Basics of Evidence-based Medicine*. BMJ Books.
- Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge University Press.
- Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge University Press.
- Gupta, A., McCalley, J., & Raghavan, B. (2020). *Emerging Technologies Beyond Moore's Law*. *IEEE Computer*, 53(1), 30-39.
- Gupta, U., Lee, Y., Choi, E., Lee, S., Verma, A., Krishna, T., ... & Ranganathan, P. (2020). *Deep Learning with Limited Numerical Precision*. *Proceedings of the 37th International Conference on Machine Learning*, 1–10.
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press.
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100.
- Harari, Y. N. (2014). *Sapiens: A Brief History of Humankind*. Harvill Secker.
- Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Spiegel & Grau.
- Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Spiegel & Grau.
- Haraway, D. (1988). *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*. Feminist Studies, 14(3), 575-599.
- Haraway, D. (1988). *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*. Feminist Studies, 14(3), 575-599.

- Hargittai, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday*, 7(4). <https://doi.org/10.5210/fm.v7i4.942> (Klasični rad o jazu u vještinama).
- Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, 11(2), 239–256.
- Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, 11(2), 239–256.
- Harries, L. (2003). *Oral Literature and Moral Values*. In *African Oral Literature: Backgrounds, Perspectives, and Continuities*, edited by D. Okpewho. Indiana University Press.
- Harris, R. (1989). *The Origin of Writing*. Open Court.
- Harris, R. (1989). *The Origin of Writing*. Open Court.
- Harris, Z. (1989). *The Evolution of Human Communication*. Harvard University Press.
- Harris, Z. (1989). *The Evolution of Human Communication*. Harvard University Press.
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (pp. 173-182). <https://doi.org/10.1145/3038912.3052569>
- Hennessy, J. L., & Patterson, D. A. (2019). *Computer Architecture: A Quantitative Approach*. Morgan Kaufmann.
- Jones, C. R., & Bergen, B. K. (2025). Large language models pass the turing test. arXiv preprint arXiv:2503.23674.
- Hennessy, J. L., & Patterson, D. A. (2019). *Computer Architecture: A Quantitative Approach* (6th ed.). Morgan Kaufmann.
- Herculano-Houzel, S. (2009). *The Human Brain in Numbers: A Linearly Scaled-up Primate Brain*. Frontiers in Human Neuroscience, 3, 31.
- Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. MIT Press.
- Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. MIT Press.
- Herculano-Houzel, S. (2016). *The Human Advantage: A New Understanding of How Our Brain Became Remarkable*. MIT Press.
- Herring, S. (2020). The coevolution of computer-mediated communication and computer-mediated discourse analysis. *Language@Internet*, 17.
- Herring, S. C. (2020). The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis. *Language@Internet*, 17.
- Hewstone, M., & Giles, H. (2018). *The Dynamics of Intergroup Communication*. Routledge.

- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250.
- Hinduja, S., & Patchin, J. W. (2015). *Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying*. Corwin Press.
- Hinduja, S., & Patchin, J. W. (2015). Bullying beyond the schoolyard: Preventing and responding to cyberbullying (2nd ed.). Corwin Press.
- Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory. Neural Computation*, 9(8), 1735–1780.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education*. Center for Curriculum Redesign.
- Hong, S., Zheng, X., Chen, J., Cheng, Y., Zhang, C., Wang, Z., ... & Zhang, A. (2023). Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*.
- Hong, S., Zheng, X., Chen, J., Cheng, Y., Zhang, C., Wang, Z., ... & Zhang, A. (2023). Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*.
- Horowitz, M. (2014). 1.1 Computing's energy problem (and what we can do about it). *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 10–14.
- Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business.
- Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Crown Business.
- Hoy, M. B. (2018). *Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants*. Medical Reference Services Quarterly, 37(1), 81–88.
- Hoy, M. B. (2018). *Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants*. Medical Reference Services Quarterly, 37(1), 81–88.
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly, 37(1), 81–88.
- <https://openai.com/index/hello-gpt-4o/>
- Huang, M. H., & Rust, R. T. (2021). *A Strategic Framework for Artificial Intelligence in Marketing*. Journal of the Academy of Marketing Science, 49(1), 30–50.
- Indiveri, G., & Liu, S. C. (2015). *Memory and information processing in neuromorphic systems*. *Proceedings of the IEEE*, 103(8), 1379–1397.
- Indiveri, G., & Liu, S. C. (2015). *Memory and Information Processing in Neuromorphic Systems*. *Proceedings of the IEEE*, 103(8), 1379–1397.
- Jenkins, H. (2008). *Social Identity*. Routledge.
- Jerison, H. J. (1973). *Evolution of the Brain and Intelligence*. Academic Press.

- Jerison, H. J. (1973). *Evolution of the Brain and Intelligence*. Academic Press.
- Jha, A., Weerkamp, W., & de Rijke, M. (2023). Moderation of Online Discussions Using Large Language Models. *Findings of the Association for Computational Linguistics: EACL 2023*, 1553-1568.
- Ji, Z., Lee, N., Fries, J., et al. (2023). Survey of Hallucination in Large Language Models. *arXiv:2302.06453*.
- Ji, Z., Lee, N., Fries, J., et al. (2023). Survey of Hallucination in Large Language Models. *arXiv: 2302.06453*.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), 1-38.
- Jia, Z., Tillman, B., Maggioni, M., & Scarpazza, D. P. (2019). *Dissecting the Graphcore IPU Architecture via Microbenchmarking*. arXiv preprint arXiv:1912.03413.
- Jiang, H., Yin, J., Cai, W., Pan, H., Liu, S., & Li, Z. (2021). *Real-time traffic signal control with adaptive deep reinforcement learning*. *IEEE Transactions on Intelligent Transportation Systems*, 22(6), 3776-3786.
- Jobin, A., Ienca, M., & Vayena, E. (2019). *The Global Landscape of AI Ethics Guidelines*. *Nature Machine Intelligence*, 1(9), 389–399.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- John Searle (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- John, A., Glendenning, A. C., Marchant, A., Montgomery, P., Stewart, A., Wood, S., ... & Hawton, K. (2018). Self-Harm, Suicidal Behaviours, and Cyberbullying in Children and Young People: Systematic Review. *Journal of Medical Internet Research*, 20(4), e129.
- Johnson, S. (2010). *Where Good Ideas Come From: The Natural History of Innovation*. Riverhead Books. (Raspravlja o važnosti slučajnosti i povezivanja ideja).
- Jouppi, N. P., et al. (2023). TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings. *Proceedings of the 50th Annual International Symposium on Computer Architecture (ISCA '23)*.
- Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., ... & Sato, K. (2017). *In-Datacenter Performance Analysis of a Tensor Processing Unit*. *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 1-12.
- Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., ... & Hogberg, D. (2017). *In-Datacenter Performance Analysis of a Tensor Processing Unit*. *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 1-12.
- Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., ... & Yoon, D. H. (2017). In-datacenter performance analysis of a tensor processing unit. *Proceedings of the 44th Annual International Symposium on Computer Architecture*, 1-12.

- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd ed.). Draft available at <https://web.stanford.edu/~jurafsky/slp3/>.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd ed.). Dostupno na <https://web.stanford.edu/~jurafsky/slp3/>
- Kassahun, A., Yu, Z., Lebedev, G., & Michelucci, D. (2021). *A Comprehensive Survey of Adaptive Learning*. *IEEE Transactions on Neural Networks and Learning Systems*, 32(9), 3715–3735.
- Kasurinen, J., & Knutas, A. (2018). Gamification in education: A review of the literature. In *Proceedings of the 22nd International Academic Mindtrek Conference* (pp. 243-250). ACM. <https://doi.org/10.1145/3275116.3275150>
- Kaye, D. B. V., Chen, X., & Zeng, J. (2022). The co-evolution of TikTok and ByteDance: Platforms, interdependencies, and global expansion. *Media, Culture & Society*, 44(3), 483–501.
- Kendall, S., & Tannen, D. (2015). Discourse and gender. *The Handbook of Discourse Analysis*, 2nd Ed., Wiley Blackwell, 639–660.
- Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019). *CTRL: A Conditional Transformer Language Model for Controllable Generation*. arXiv preprint arXiv:1909.05858.
- Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019). *CTRL: A Conditional Transformer Language Model for Controllable Generation*. arXiv preprint arXiv:1909.05858.
- Khan Academy. (2024). Khanmigo - AI Guide. Khan Academy. <https://www.khanacademy.org/khan-labs>
- Khan Academy. (2024). *Khanmigo - AI Guide*. Khan Academy. <https://www.khanacademy.org/khan-labs>
- Klarna. (2024, February 27). Klarna AI assistant handles two-thirds of customer service chats in its first month. Klarna Newsroom. <https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/>
- Klarna. (2024, February 27). *Klarna AI assistant handles two-thirds of customer service chats in its first month*. Klarna Newsroom. <https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/>
- Klein, R. G. (2009). *The Human Career: Human Biological and Cultural Origins*. University of Chicago Press.
- Kövecses, Z. (2010). *Metaphor: A Practical Introduction*. Oxford University Press.

- Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). *Cyberbullying: Bullying in the Digital Age*. Wiley-Blackwell.
- Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). *Cyberbullying: Bullying in the digital age* (2nd ed.). Wiley-Blackwell.
- Kramer, S. N. (1963). *History Begins at Sumer*. University of Pennsylvania Press.
- Kramer, S. N. (1963). *History Begins at Sumer*. University of Pennsylvania Press.
- Kramer, S. N. (1963). *The Sumerians: Their History, Culture, and Character*. University of Chicago Press.
- Kramer, S. N. (1963). *The Sumerians: Their History, Culture, and Character*. University of Chicago Press.
- Kraus, S. (1960). *The Great Debates: Background, Perspective, Effects*. Indiana University Press.
- Kraus, S. (1960). *The Great Debates: Kennedy vs. Nixon, 1960*. Indiana University Press.
- Kshetri, N. (2023). Ransomware economy: ecosystem, actors, processes, market dynamics, and economic impacts. *Journal of Cybersecurity*, 9(1), tyad001. <https://doi.org/10.1093/cybsec/tyad001>
- Kuhail, M. A., Menezes, R., & Farooq, S. (2023). Conversational AI in Education. *ACM Transactions on Learning Technologies*, 16(4), 1–24. <https://doi.org/10.1145/3608488>
- Kuhlmann, I., Piktus, A., Pyatkin, V., et al. (2022). A Mixed-Methods Approach to Characterizing Biases in Large Language Models. *ACL 2022*.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- Kumar, P., & Chattopadhyay, A. (2019). *Beyond Moore's Law: The Growing Gap Between Processor and Memory Technology*. *Proceedings of the IEEE*, 107(1), 2-4.
- Kumar, S., & Chattopadhyay, A. (2019). *Beyond Moore's Law: Opportunities in Terahertz Science and Technology*. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(4), 1–13.
- Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking.
- Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Viking.
- Kurzweil, R. (2005). The Singularity Is Near: When Humans Transcend Biology. Viking.
- Kurzweil, R. (2005). The Singularity Is Near: When Humans Transcend Biology. Viking.
- Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.
- Labov, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Blackwell.
- Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). *A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics*. *AI Magazine*, 38(4), 13–26.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). *Building Machines That Learn and Think Like People*. *Behavioral and Brain Sciences*, 40.

- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. The University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- LangChain. (2024). *LangGraph: Building complex, stateful agent applications*.  
LangChain Documentation. <https://python.langchain.com/docs/langgraph>
- Laudon, K. C., & Traver, C. G. (2016). *E-commerce 2016: Business, Technology, Society*. Pearson.
- Laudon, K. C., & Traver, C. G. (2016). *E-commerce 2016: Business, Technology, Society*. Pearson.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. In Pervin, L. A. (Ed.), *Handbook of Personality: Theory and Research* (pp. 609-637). Guilford Press.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. In Pervin, L. A. (Ed.), *Handbook of Personality: Theory and Research* (pp. 609-637). Guilford Press.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, 521(7553), 436–444.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, 521(7553), 436-444.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, 521(7553), 436–444.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. *Nature*, 521(7553), 436-444.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lee, J. D., Reddy, S., Jin, H., Shi, A., Lee, K., Ponce Wong, N., ... & Zhang, H. (2022).  
*CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities*. arXiv preprint arXiv:2201.06796.
- Lee, J. D., Reddy, S., Jin, H., Shi, A., Lee, K., Ponce Wong, N., ... & Zhang, H. (2022).  
*CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities*. arXiv preprint arXiv:2201.06796.
- Lee, J. D., Reddy, S., Jin, H., Shi, A., Lee, K., Ponce Wong, N., ... & Zhang, H. (2022).  
*CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities*. arXiv preprint arXiv:2201.06796.
- Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2021). Deduplicating Training Data Makes Language Models Better. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8424–8445.
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (2009). *A Brief History of the Internet*. ACM SIGCOMM Computer Communication Review, 39(5), 22–31.
- Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.
- Levinson, S. C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.

- Levy, A. (1999). *Orality and Identity: The Cultural Constructs of Oral Societies*. *Journal of Cultural Studies*, 14(2), 45–63.
- Lewis, C. (2020). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. *Advances in Neural Information Processing Systems*, 33, 9459-9474. (Originalni RAG rad).
- Lewis, T. (1991). *Empire of the Air: The Men Who Made Radio*. HarperCollins.
- Lewis, T. (1991). *Empire of the Air: The Men Who Made Radio*. HarperCollins.
- Li, C., Nikolakopoulos, A. N., Shang, M., Ren, Z., Yan, H., Naidu, D., ... & Bär, D. (2023a). Large Language Models for Generative Recommendation: A Survey and Visionary Discussions. *arXiv preprint arXiv:2309.01157*. <https://arxiv.org/abs/2309.01157>
- Li, M., Huo, Z., & Chen, M. (2023b). The impact of AI-powered personalization on customer engagement and purchase intention in e-commerce live streaming. *Journal of Research in Interactive Marketing*, 17(4), 634-652. <https://doi.org/10.1108/JRIM-12-2022-0388>
- Li, S., & Deng, W. (2020). *Deep Facial Expression Recognition: A Survey*. *IEEE Transactions on Affective Computing*, 1–1.
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., ... & Koreeda, Y. (2021). *Holistic Evaluation of Language Models*. Stanford Institute for Human-Centered AI. [Potražiti najnoviju verziju HELM izvještaja]
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., ... & Koreeda, Y. (2021). Report from the Workshop on Foundation Models. Stanford Institute for Human-Centered Artificial Intelligence (HAI). <https://cfrfm.stanford.edu/report.html>
- Liang, W., Al-Hazmi, A., Iyyer, M., & Bansal, M. (2023). Can Large Language Models Be Good Companions for Exploratory Data Analysis? Findings of the Association for Computational Linguistics: EMNLP 2023, 1405–1425. <https://aclanthology.org/2023.findings-emnlp.95>
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80. (Klasični rad o jednom od najpoznatijih sustava preporuka).
- Lipartito, K. (2003). *Picturephone and the Information Age: The Social Meaning of Failure*. *Technology and Culture*, 44(1), 50–81.
- Lipartito, K. (2003). *The Bell System and Regional Business Elites*. The Johns Hopkins University Press.
- Lipartito, K. (2003). *The Bell System and Regional Business Elites*. The Johns Hopkins University Press.

- Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. *LREC 2016*.
- Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. *LREC 2016*.
- Litman, T. (2020). *Autonomous Vehicle Implementation Predictions: Implications for Transport Planning*. Victoria Transport Policy Institute.
- Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692.
- Lucy, L., & Bamman, D. (2021). Gender and Representation Bias in GPT-3 Generated Stories. *ACL 2021*.
- Luhmann, N. (2004). *Social Systems*. Stanford University Press.
- Malik, A., Budhwar, P., & Kazmi, B. A. (2023). Artificial intelligence (AI)-assisted HRM: Towards an extended strategic framework. *Human Resource Management Journal*, 33(4), 870-893. <https://doi.org/10.1111/1748-8583.12514>
- Mamykina, L., Manoim, B., Mittal, M., Hripcak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. *CHI 2011*.
- Mamykina, L., Manoim, B., Mittal, M., Hripcak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. *CHI 2011*.
- Mamykina, L., Manoim, B., Mittal, M., Hripcak, G., & Hartmann, B. (2011). Design lessons from the fastest q&a site in the west. *Proceedings of the SIGCHI conference on human factors in computing systems*, 2857-2866.
- Manovich, L. (2001). *The Language of New Media*. MIT Press.
- Manovich, L. (2023). *AI Aesthetics*. Strelka Press. [Ili noviji rad Manovicha ako postoji]
- Marconi, G. (1902). *Syntonic Wireless Telegraphy*. The Electrician Printing and Publishing Company.
- Marconi, G. (1902). *Wireless Telegraphic Communication*. Nobel Lecture.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books.
- Markram, H., Kuhn, R., Esteban, J. A., Schürmann, F., Andersen, N., Muller, E., ... & Ramaswami, M. (2015). Reconciling the sparse connectivity of neocortical networks with integrative dendritic computation. *Neuron*, 86(2), 369-381.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., ... & Schürmann, F. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2), 456-492.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., ... & Schürmann, F. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2), 456-492.
- Marsh, D. (1991). *Television: Critical Methods and Applications*. Routledge.

- Marvin, C. (1988). *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*. Oxford University Press.
- Marwick, A. E., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*. Data & Society Research Institute.
- Marwick, A. E., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute.
- Marwick, A. E., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>
- Marwick, A. E., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute.
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online*. Data & Society Research Institute.
- Massanari, A. L. (2015). *Participatory Culture, Community, and Play: Learning from Reddit*. Peter Lang.
- Massanari, A. L. (2017). # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New media & society*, 19(3), 329-346. (Primjer analize kulture specifične platforme).
- Mateo, A. G., Ruiz, F. J., Perez-Sanagustin, M., Broisin, J., Fonseca, D., & Munoz-Organero, M. (2024). Large language models as brainstorming assistants in conceptual design. *Computers & Education: Artificial Intelligence*, 7, 100081. <https://doi.org/10.1016/j.caeari.2024.100081>
- Matsumoto, D., & Hwang, H. S. (2012). *Culture and Emotion: The Integration of Biological and Cultural Contributions*. Journal of Cross-Cultural Psychology, 43(1), 91-118.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. *ACL 2020*.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. *ACL 2020*.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. **ACL 2020**.
- Mazzone, M., & Elgammal, A. (2019). Art, creativity, and the potential of artificial intelligence. *Arts*, 8(1), 26.
- McDuff, D., El Kaliouby, R., Cohn, J. F., & Picard, R. W. (2015). *Predicting Ad Liking and Purchase Intent: Large-scale Analysis of Facial Responses to Ads*. IEEE Transactions on Affective Computing, 6(3), 223–235.
- McGregor, J., & Holmes, J. (2020). Leadership, discourse, and ethnicity. *The Routledge Handbook of Language in the Workplace*, Routledge, 406–419.
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.

- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35.
- Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G., & Cubuk, E. D. (2023). Scaling deep learning for materials discovery. *Nature*, 619(7970), 518-525.
- Meta AI. (2023). LLaMA 2. Preuzeto s: <https://ai.meta.com/llama/>
- Meta AI. (2023). LLaMA 2. Preuzeto s: <https://ai.meta.com/llama/>
- Meta AI. (2023). LLaMA 2. Preuzeto s: <https://ai.meta.com/llama/>
- Meta AI. (2023, May 19). *Meta's next-generation AI infrastructure*. Meta AI Blog. <https://ai.meta.com/blog/meta-next-generation-ai-infra/>
- Meta AI. (2024, April 18). *Introducing Meta Llama 3: The most capable openly available LLM to date*. Meta AI Blog. <https://ai.meta.com/blog/meta-llama-3/>
- Microsoft Azure. (2024). *Azure AI Infrastructure*. [Potražiti najnoviju dokumentaciju o Azure ND & NC seriji VM-ova i AI superračunalima]
- Microsoft. (2024). Microsoft Copilot. Microsoft. [Potražiti najnovije informacije o Copilot integracijama]
- Mijwil, M. M., Aljanabi, M., Ali, A. H., & Abttan, R. A. (2023). Chatgpt: Exploring the Role of Cybersecurity in the Protection of Large Language Models (LLMs). *Mesopotamian Journal of Cybersecurity*, 2023, 54-58. [Napomena: Iako naslov sugerira cybersecurity, rad često diskutira i općenite interne primjene]
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *NeurIPS 2013*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint arXiv:1301.3781.
- Miller, E. K., & Cohen, J. D. (2001). *An Integrative Theory of Prefrontal Cortex Function*. *Annual Review of Neuroscience*, 24(1), 167–202.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1), 167-202.
- Mishra, S., Dhamala, J., Arora, S., et al. (2022). Improving Factuality and Reasoning in Language Models through Multi-Step Reasoning and Debiasing. *EMNLP 2022*.
- Mistral AI. (2023). Mistral Model Card. Preuzeto s: <https://mistral.ai/>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). *Model Cards for Model Reporting*. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2019). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679. [Original je iz 2016, ali citiran i relevantan i kasnije]

- Moore, G. E. (1965). *Cramming more components onto integrated circuits*. *Electronics*, 38(8).
- Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8), 114-117.
- Mulki, J. P., Bardhi, F., Lassk, F. G., & Nanavaty-Dahl, J. (2009). Set up remote workers to thrive. *MIT Sloan Management Review*, 51(1), 63–69.
- Mulki, J. P., Bardhi, F., Lassk, F. G., & Nanavaty-Dahl, J. (2009). *Set Up Remote Workers to Thrive*. MIT Sloan Management Review, 51(1), 63–69.
- Müller, V. C., & Bostrom, N. (2016). *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*. U *Fundamental Issues of Artificial Intelligence* (str. 555–572). Springer.
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103.
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103.
- Needham, J. (1985). *Science and Civilisation in China. Vol. 5: Chemistry and Chemical Technology*. Cambridge University Press.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2023). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 4, 100108.
- Niane, D. T. (1960). *Sundiata: An Epic of Old Mali*. Longmans.
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119-158.
- Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press. (Iako iz 2009., ostaje ključni rad o kontekstualnoj privatnosti).
- Nkambou, R., Bourdeau, J., & Mizoguchi, R. (Eds.). (2010). *Advances in intelligent tutoring systems*. Springer.
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- Nov, O., Shabat, G., Gordon, S., Friedberg, G., Zucker, L., Avidan, Y., ... & Reis, B. Y. (2024). Evaluating large language model-based applications for triage using emergency severity index vignettes. *npj Digital Medicine*, 7(1), 79. <https://doi.org/10.1038/s41746-024-01063-x>
- NVIDIA. (2020). *NVIDIA A100 Tensor Core GPU Architecture*. NVIDIA Corporation.
- NVIDIA. (2022, March 22). *NVIDIA Hopper Architecture In-Depth*. NVIDIA Technical Blog. <https://developer.nvidia.com/blog/nvidia-hopper-architecture-in-depth/>
- NVIDIA. (2023, November 13). *NVIDIA HGX H200 Platform*. NVIDIA. <https://www.nvidia.com/en-us/data-center/hgx-h200/>

- NVIDIA. (2024, March 18). *NVIDIA Blackwell Platform Arrives to Power a New Era of Computing*. NVIDIA Newsroom. <https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing>
- Ogata, T., Okazaki, N., Otake, N., et al. (2019). Towards the Next Generation of Multilingual Call Centers: A Study on Cultural Sensitivity in Conversational AI. *PACLIC 2019*.
- Ogata, T., Okazaki, N., Otake, N., et al. (2019). Towards the Next Generation of Multilingual Call Centers: A Study on Cultural Sensitivity in Conversational AI. *PACLIC 2019*.
- Ogden, C. K., & Richards, I. A. (1923). *The Meaning of Meaning*. Harcourt, Brace & Company.
- Ogden, C. K., & Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Kegan Paul, Trench, Trubner & Co.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.
- Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge.
- Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the Word*. Routledge.
- OpenAI. (2023). GPT-4 Technical Report. <https://openai.com/research/gpt-4>
- OpenAI. (2023). *GPT-4 Technical Report*. Preuzeto s <https://openai.com/research/gpt-4>
- OpenAI. (2023). *GPT-4 Technical Report*. arXiv preprint arXiv:2303.08774. (Uključuje i sekcije o rizicima i mitigaciji).
- OpenAI. (2023). GPT-4 Technical Report. arXiv preprint arXiv:2303.08774. <https://arxiv.org/abs/2303.08774>
- OpenAI. (2023, March 14). *GPT-4*. OpenAI Blog. <https://openai.com/research/gpt-4>
- OpenAI. (2023, March 14). *GPT-4*. OpenAI Research. <https://openai.com/research/gpt-4> [Posebno vidjeti primjere primjene u obrazovanju]
- OpenAI. (2023b). GPT-4o. Preuzeto s: <https://openai.com/>
- OpenAI. (2024a). *GPT-4 Turbo*.
- OpenAI. (2024b, May 13). Introducing GPT-4o. OpenAI Blog.
- OpenAI. (2024b, May 13). Introducing GPT-4o. OpenAI Blog. <https://openai.com/index/hello-gpt-4o/>
- Pagels, A. E. (1979). *The Alexandrian Library in the Roman World*. Harvard University Press.
- Pagels, E. (1979). *The Gnostic Gospels*. Random House.
- Pagels, E. (1979). *The Gnostic Gospels*. Random House.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.

- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.
- Parkinson, R., & Quirke, S. (1995). *Papyrus*. British Museum Press.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press. (Istražuje utjecaj skrivenih algoritama).
- Patchin, J. W., & Hinduja, S. (2010). Cyberbullying and self-esteem. *Journal of School Health*, 80(12), 614-621.
- Patchin, J. W., & Hinduja, S. (2010). Cyberbullying and self-esteem. *Journal of school health*, 80(12), 614-621.
- Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., ... & Dean, J. (2022). *Carbon Emissions and Large Neural Network Training*. arXiv preprint arXiv:2104.10350.
- Pearce, H., Tan, B., Ahmad, B., Karamongkol, R., & Raman, D. (2022). Asleep at the Keyboard? Assessing the Security of GitHub Copilot's Code Contributions. Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP).
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *EMNLP 2014*.
- Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global Vectors for Word Representation*. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1532–1543.
- Pentland, A. (2014). *Social Physics: How Good Ideas Spread—The Lessons from a New Science*. Penguin Press. (Iako starija, relevantna za analizu društvenih interakcija putem podataka).
- Perez, E., Chen, L., Kasai, J., et al. (2022). Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned. *NeurIPS 2022*.
- Peters, O. (2003). *Learning with New Media in Distance Education*. Kogan Page.
- Peters, O. (2003). *Learning with New Media in Distance Education*. Kogan Page.
- Picard, R. W. (1997). *Affective Computing*. MIT Press.
- Picard, R. W. (1997). *Affective Computing*. MIT Press.
- Pinker, S. (1994). *The Language Instinct*. William Morrow.
- Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. William Morrow and Company.
- Pop, E., Sinha, S., & Goodson, K. E. (2010). *Heat Generation and Transport in Nanometer-Scale Transistors*. Proceedings of the IEEE, 94(8), 1587–1601.
- Prabhakaran, V., coarse, S., & Gebru, T. (2022). Whose Culture? Towards a Systemic Understanding of Culturally-Situated NLP. Proceedings of the 2022 Conference on Fairness, Accountability, and Transparency (FAccT '22), 690-703.

- Przybylski, A. K., & Weinstein, N. (2013). Can you connect with Tom? An exploration of parasocial interaction alignment. *Cyberpsychology, Behavior, and Social Networking*, 16(7), 481-487. <https://doi.org/10.1089/cyber.2012.0339>
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2022). *Pre-trained Models for Natural Language Processing: A Survey*. *Science China Technological Sciences*, 65(1), 1–26.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2022). *Pre-trained Models for Natural Language Processing: A Survey*. *Science China Technological Sciences*, 65(1), 1–26.
- Quan-Haase, A., & Wellman, B. (2002). How does the Internet affect social capital. In *Social capital and information technology* (pp. 113–135). MIT Press. (Rani rad Wellmana i suradnika o vezi online-offline).
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI Blog, 1(8), 9.
- Radziwill, N. M., & Benton, M. C. (2017). *Evaluating Quality of Chatbots and Intelligent Conversational Agents*. *Software Quality Professional*, 19(3), 25–36.
- Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Sifre, L., ... & Irving, G. (2021). *Scaling Language Models: Methods, Analysis & Insights from Training Gopher*. arXiv preprint arXiv:2112.11446.
- Rae, J., Borgeaud, S., Cai, T., et al. (2021). Scaling Language Models: Methods, Analysis & Insights from Training Gopher. *DeepMind Technical Report*.
- Rae, J., Borgeaud, S., Cai, T., et al. (2021). Scaling Language Models: Methods, Analysis & Insights from Training Gopher. *DeepMind Technical Report*.
- Raffel, C. et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *JMLR*, 21(140).
- Rainie, L., & Wellman, B. (2012). *Networked: The New Social Operating System*. MIT Press.
- Rainie, L., & Wellman, B. (2012). *Networked: The New Social Operating System*. MIT Press. (Sveobuhvatan pregleđ umreženog individualizma).
- Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020). *Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing*. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 145-151.
- Raymond, E. S. (1999). *The Cathedral & the Bazaar*. O'Reilly Media.
- Raymond, E. S. (1999). *The Cathedral & the Bazaar*. O'Reilly Media.
- Raymond, E. S. (1999). *The Cathedral and the Bazaar*. O'Reilly Media.
- Reed, J., & Dongarra, J. (2015). *Exascale computing and big data*. *Communications of the ACM*, 58(7), 56–68.
- Reepa, B. (2017). Evolving Cybersecurity Landscapes. *IT Professional*, 19(4), 16-20.

- Regier, T., & Kay, P. (2009). *Color Categories are not Universal: Replications and Extensions of Berlin and Kay (1969)*. Journal of Experimental Psychology: General, 138(3), 329-344.
- Regier, T., & Kay, P. (2009). *Color Categories are not Universal: Replications and Extensions of Berlin and Kay (1969)*. Journal of Experimental Psychology: General, 138(3), 329-344.
- Regier, T., & Kay, P. (2009). *Color Categories are not Universal: Replications and Extensions of Berlin and Kay (1969)*. Journal of Experimental Psychology: General, 138(3), 329-344.
- Rehm, G., & Uszkoreit, H. (Eds.). (2013). *META-NET White Papers Series: Europe's Languages in the Digital Age*. Springer.
- Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley.
- Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley.
- Ribeiro, M. T., Wu, T., Guestrin, C., & Singh, S. (2020). Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. *ACL 2020*.
- Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. In *Recommender Systems Handbook* (pp. 1-35). Springer US. (Pregled područja sustava preporuka).
- Rifkin, J. (2014). *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*. Palgrave Macmillan. (Istražuje ekonomski i društvene posljedice automatizacije i Interneta stvari).
- Roberts, S. T. (2019). *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.
- Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., ... & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, 18(5), 569-586.
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *TACL 2020*.
- Roper, G. (2017). *Arabic Printing in the Islamic World*. In J. Daybell & A. Gordon (Eds.), *Cultures of Correspondence in Early Modern Britain*. University of Pennsylvania Press.
- Rosa, J., & Flores, N. (2017). Unsettling Race and Language: Toward a Raciolinguistic Perspective. *Language in Society*, 46(5), 621-647.
- Rosen, D., Lafontaine, P. R., & Hendrickson, B. (2011). CouchSurfing: Belonging and trust in a globally cooperative online social network. *New Media & Society*, 13(6), 981-998.
- Roy, G., Khare, A., Datta, B., & Sivakumar, T. C. (2024). Impact of conversational AI on customer engagement and loyalty: A study of the banking industry. *International Journal of Information Management*, 76, 102762. <https://doi.org/10.1016/j.ijinfomgt.2024.102762>

- Rubin, D. C. (1995). *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes*. Oxford University Press.
- Ruder, S., Cotterell, R., Kementchedjhieva, Y., & Kumar, S. (2019). A Primer in Bilingual Lexicon Induction. *JAIR*, 65, 681–743.
- Ruder, S., Vulic, I., & Søgaard, A. (2019). *A Survey of Cross-lingual Word Embedding Models*. *Journal of Artificial Intelligence Research*, 65, 569–630.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4. izd.). Pearson.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.
- Rust, R. T., & Huang, M. H. (2024). The feeling economy: Managing customer emotions with artificial intelligence. *California Management Review*, 66(2), 7-30. <https://doi.org/10.1177/00081256231214521>
- Sadiku, M. N., Adebiyi, A. A., & Musa, S. M. (2021). Ambient intelligence: A primer. *International Journal of Advanced Research in Computer Science and Software Engineering*, 11(3), 1-6.
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2024). Artificial intelligence challenges in scientific writing: delving into the shady side of ChatGPT. *Critical Care*, 28(1), 1-3.
- Saussure, F. (1916). *Course in General Linguistics*. McGraw-Hill.
- Saussure, F. (1916). *Course in General Linguistics*. McGraw-Hill.
- Saussure, F. de. (1916). *Course in General Linguistics*. McGraw-Hill.
- Saussure, F. de. (1916). *Course in General Linguistics*. McGraw-Hill.
- Saussure, F. de. (1916). *Course in General Linguistics*. McGraw-Hill.
- Saussure, F. de. (1916). *Course in General Linguistics*. McGraw-Hill.
- Schaller, R. R. (1997). *Moore's law: past, present and future*. *IEEE Spectrum*, 34(6), 52–59.
- Schudson, M. (1993). *Advertising, The Uneasy Persuasion: Its Dubious Impact on American Society*. Basic Books.
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). *Green AI*. *Communications of the ACM*, 63(12), 54–63.
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Searle, J. R. (1995). *The Construction of Social Reality*. Free Press.

- Seering, J., Kraut, R. E., & Kaufman, G. (2019). Supporting newcomer socialization and engagement through structured introductions. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-23. (Jedan od radova Seeringa o Discordu, iako fokusiran na pridošlice).
- Seering, J., Kraut, R. E., & Kaufman, G. (2020). Beyond the Talking Stage: The Role of Moderator Feedback in Reddit Commenting Behavior. *CSCW '20*.
- Seering, J., Luria, M., Kaufman, G., & Hammer, J. (2020). Beyond dyadic interactions: Considering chatbots as community members. *CSCW 2020*.
- Settles, B., Lachlan-Hachey, E., Peterson, C., Brust, C., Angelico, R., Hargett, B., ... & Bicknell, K. (2018). *Birdbrain: A Framework for Improving Duolingo Learning Outcomes*. Duolingo Research. [Potražiti noviju dokumentaciju ako postoji, ovo je raniji rad na njihovom AI sustavu]
- Shamekhi, A., Lorenz, P., Lam, S., & Kloeckner, K. (2023). *Personalized Large Language Models*. arXiv preprint arXiv:2311.16010.
- Shamekhi, A., Lorenz, P., Lam, S., & Kloeckner, K. (2023). *Personalized Large Language Models*. arXiv preprint arXiv:2311.16010.
- Shamekhi, A., Lorenz, P., Lam, S., & Kloeckner, K. (2023). *Personalized Large Language Models*. arXiv preprint arXiv:2311.16010. <https://arxiv.org/abs/2311.16010>
- Shannon, C. E. (1948). *A Mathematical Theory of Communication*. *Bell System Technical Journal*, 27(3), 379–423.
- Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.
- Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press.
- Shen, C., Nguyen, D., Yan, R., ... & Rajpurkar, P. (2024). Large language models are clinical multitask learners. *Nature*, 629(8011), 405-412. <https://doi.org/10.1038/s41586-024-07256-2>
- Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., ... & Soljačić, M. (2017). *Deep learning with coherent nanophotonic circuits*. *Nature Photonics*, 11(7), 441–446.
- Sheng, E., Chang, K. W., & Stroulia, E. (2021). *Societal Biases in Language Generation: Progress and Challenges*. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, 4275-4293.
- Sheng, E., Chang, K. W., Natarajan, P., & Peng, N. (2021). Societal Biases in Language Generation: Progress and Challenges. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorials*, 16-21.
- Shin, R., et al. (2020). Coherent Dialogue Generation from Structured Text. *EMNLP 2020*.
- Shin, R., Ribeiro, M. T., Ghassemi, M., & Beaugureau, J. (2020). *On the Pitfalls of Analyzing Individual Neurons in Language Models*. arXiv preprint arXiv:2012.14028.

- Shin, R., Ribeiro, M. T., Ghassemi, M., & Beaugureau, J. (2020). *On the Pitfalls of Analyzing Individual Neurons in Language Models*. arXiv preprint arXiv:2012.14028.
- Shum, H. Y., He, X., & Li, D. (2018). *From Eliza to Xiaolce: Challenges and Opportunities with Social Chatbots*. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10–26.
- Shum, H.-Y., He, X., & Li, D. (2018). From Eliza to Xiaolce: Challenges and Opportunities with Social Chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10–26.
- Siddiq, M. L., Santos, J., Santos, R., Lopes, P. P., & Santos, N. (2024). Demystifying Exploitable Vulnerabilities in Large Language Model-Based Applications. Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24). [Provjeriti točan status publikacije ako je još preprint]
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). *Mastering the Game of Go with Deep Neural Networks and Tree Search*. *Nature*, 529(7587), 484–489.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). *Mastering the game of Go without human knowledge*. *Nature*, 550(7676), 354–359.
- Slapper, G., & Kelly, D. (2015). *The English Legal System*. Routledge.
- Smith, A. D. (1991). *National Identity*. University of Nevada Press.
- Smith, B. (2006). *Publishing Research Results*. *Radiology*, 238(1), 3–4.
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry*, 49(4), 376–385.
- Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of child psychology and psychiatry*, 49(4), 376–385.
- Solove, D. J. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York University Press.
- Solove, D. J. (2004). *The Digital Person: Technology and Privacy in the Information Age*. NYU Press.
- Solove, D. J. (2004). The digital person: Technology and privacy in the information age. NYU press.
- Spigel, L. (1992). *Make Room for TV: Television and the Family Ideal in Postwar America*. University of Chicago Press.
- Spigel, L. (1992). *Make Room for TV: Television and the Family Ideal in Postwar America*. University of Chicago Press.
- Sporns, O. (2013). Network analysis, complexity, and brain function. *Complexity*, 18(5), 8–15.

- Sporns, O. (2013). *Structure and Function of Complex Brain Networks*. Dialogues in Clinical Neuroscience, 15(3), 247–262.
- Sporns, O. (2018). *Graph Theory Methods: Applications in Brain Networks*. Dialogues in Clinical Neuroscience, 20(2), 111–121.
- Sporns, O. (2018). *Networks of the Brain*. MIT Press.
- Staal, F. (1986). *The fidelity of oral tradition and the origins of science*. North-Holland Publishing Company. (Ključni rad o vjernosti vedske usmene predaje).
- Standage, T. (1998). *The Victorian Internet*. Walker & Company.
- Standage, T. (1998). *The Victorian Internet*. Walker & Company.
- Standage, T. (1998). *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers*. Walker and Company.
- Standage, T. (1998). *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers*. Walker Publishing Company.
- Standage, T. (1998). *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's Online Pioneers*. Walker Publishing Company.
- Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). *Social media analytics—Challenges in topic discovery, data collection, and data preparation*. International Journal of Information Management, 39, 156–168.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. arXiv preprint arXiv:1906.02243.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.
- Sundararajan, A. (2016). *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*. MIT Press.
- Sundararajan, A. (2016). *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*. MIT Press.
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2).
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1410>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT press.

- Tattersall, I. (2017). *How Can We Detect When Language Emerged?*. Psychonomic Bulletin & Review, 24(1), 64–67.
- Buckner, R. L., & Krienen, F. M. (2013). *The Evolution of Human Brain Organization: Convergent Evidence from Comparative Studies*. *Frontiers in Human Neuroscience*, 7, 111.
- Terblanche, N. (2023). The role of conversational artificial intelligence in customer service: A systematic literature review and future research agenda. *Journal of Service Theory and Practice*, 33(5), 684-710. <https://doi.org/10.1108/JSTP-07-2022-0155>
- Theis, T. N., & Wong, H. S. P. (2017). *The End of Moore's Law: A New Beginning for Information Technology*. *Computing in Science & Engineering*, 19(2), 41–50.
- Theis, T. N., & Wong, H. S. P. (2017). The end of Moore's law: A new beginning for information technology?. *Computing in Science & Engineering*, 19(2), 41-50.
- thumb\_upthumb\_down33.8s
- TII. (2023). Falcon Language Model. Preuzeto s: <https://falconllm.tii.ae/>
- Timms, M. J. (2016). Technology enhanced assessment. In *Handbook of Research on Learning and Instruction* (pp. 561-578). Routledge.
- Topol, E. J. (2019). Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books.
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. *arXiv preprint arXiv:2307.09288*.
- Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., ... & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Hewlett Foundation Report*.
- Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.
- Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.
- Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.
- Tufekci, Z. (2017). Twitter and Tear Gas: The Power and Fragility of Networked Protest. Yale University Press.
- Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
- Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
- Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
- Turkle, S. (2011). Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books.
- Turkle, S. (2017). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.

- Turkle, S. (2017). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books. (Klasik o utjecaju tehnologije na meduljudske odnose).
- Twenge, J. M. (2019). The sad state of happiness in the United States and the role of digital media. *World Happiness Report 2019*, 86-99.
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Pariz.
- Vaidyam, A. N., Wisniewski, H., Halama, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Canadian Journal of Psychiatry*, 64(7), 456–464.
- van Dijk, J. (2005). *The Deepening Divide: Inequality in the Information Society*. SAGE Publications.
- van Dijk, J. (2005). *The Deepening Divide: Inequality in the Information Society*. SAGE Publications.
- van Dijk, J. (2005). *The Deepening Divide: Inequality in the Information Society*. Sage Publications.
- van Dijk, J. A. G. M. (2020). *The Digital Divide*. Polity Press. (Novije izdanje ključnog rada).
- Van Esch, P., Cui, Y., & Jain, V. (2021). Lead generation using chatbots: Strategizing the use of AI. *Journal of Retailing and Consumer Services*, 60, 102481. <https://doi.org/10.1016/j.jretconser.2021.102481>
- VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197-221.
- VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197-221. <https://doi.org/10.1080/00461520.2011.611369>
- Vansina, J. (1985). *Oral Tradition as History*. University of Wisconsin Press.
- Vaswani, A. et al. (2017). Attention Is All You Need. *NeurIPS 2017*.
- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. *NeurIPS 2017*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is All You Need. Advances in Neural Information Processing Systems*, 30.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is All You Need. Advances in Neural Information Processing Systems*, 30, 5998–6008.
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 233.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

- Vraga, E. K., & Tully, M. (2019). News Literacy, Social Media Behaviors, and Political Attitudes: Assessing Effects on Exposure, Knowledge, and Engagement. *Digital Journalism*, 7(8), 1052–1071.
- Vraga, E. K., Tully, M., & Rojas, H. (2020). Media literacy messages and hostile media perceptions: Processing of nonpartisan versus partisan political information. *Mass Communication and Society*, 23(6), 876-901. (Primjer istraživanja o medijskoj pismenosti).
- Vygotsky, L. S. (1986). *Thought and Language*. MIT Press.
- Vygotsky, L. S. (1986). *Thought and Language*. MIT Press.
- Vygotsky, L. S. (1986). *Thought and Language*. MIT Press.
- Vygotsky, L. S. (1986). *Thought and Language*. MIT Press.
- Waldrop, M. M. (2016). The chips are down for Moore's law. *Nature News*, 530(7589), 144.
- Waldrop, M. M. (2016). The chips are down for Moore's law. *Nature News*, 530(7589), 144.
- Wallace, R. S. (2003). *The Anatomy of A.L.I.C.E.*. In *Parsing the Turing Test* (pp. 181-210). Springer, Dordrecht. (Opisuje AIML i A.L.I.C.E.).
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... & Wang, L. (2023). A survey on large language model based autonomous agents. *arXiv preprint arXiv:2308.11432*.
- Ward, N., & De Vries, B. (2021). Nonverbal Communication and Conversational Agents. *ACM Transactions on Interactive Intelligent Systems*.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe report DGI(2017)09. <https://rm.coe.int/information-disorder-report-november-2017/1680764666>
- Wardle, C., & Derakhshan, H. (2017). Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report DGI(2017)09.
- Weaviate Blog. (Accessed 2024). [Potražiti relevantne postove, npr. "Advanced RAG techniques", "Hybrid Search", "Improving RAG performance"]. Weaviate.io. (Weaviate često objavljuje kvalitetne članke o RAG-u i vektorskim bazama).
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), 214-229.

- Weidinger, L., Uesato, J., Rauh, M., et al. (2022). Ethical and Social Risks of Harm from Language Models. *ACL 2022*.
- Weidinger, L., Uesato, J., Rauh, M., et al. (2022). Ethical and social risks of harm from Language Models. *ACL 2022*.
- Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, 265(3), 94-104.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. (Originalni rad o ELIZA-i).
- Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 25(2), 227-252.
- Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 25(2), 227-252.
- Wellman, B., & Haythornthwaite, C. (2002). *The Internet in Everyday Life*. Blackwell.
- Wellman, B., & Haythornthwaite, C. (Eds.). (2002). *The Internet in Everyday Life*. Blackwell Publishers.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39-52.
- Wheeler, T. (2000). *Mr. Lincoln's T-Mails: How Abraham Lincoln Used the Telegraph to Win the Civil War*. HarperCollins.
- Wheeler, T. (2000). *Mr. Lincoln's T-Mails: How Abraham Lincoln Used the Telegraph to Win the Civil War*. HarperCollins.
- Wheeler, T. (2000). *Mr. Lincoln's T-Mails: The Untold Story of How Abraham Lincoln Used the Telegraph to Win the Civil War*. HarperCollins.
- Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings*. MIT Press.
- Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press.
- Wikimedia Foundation. (2023, November 1). Testing New Enterprise Tools to Support Wikimedia Content. Wikimedia Diff. <https://diff.wikimedia.org/2023/11/01/testing-new-enterprise-tools-to-support-wikimedia-content/>
- Williams, R. (1974). *Television: Technology and Cultural Form*. Fontana Press.
- Williams, R. (1974). *Television: Technology and Cultural Form*. Schocken Books.
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114-123.
- Winawer, J., Witthoft, N., Frank, M. C., & Viola, F. (2007). *Russian and English experience color in different ways*. Proceedings of the National Academy of Sciences, 104(19), 7780-7785.

- Winawer, J., Witthoft, N., Frank, M. C., et al. (2007). Russian blues reveal effects of language on color discrimination. *PNAS*, 104(19), 7780–7785.
- Winkler, R., & Söllner, M. (2018). *Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis*. Academy of Management Annual Meeting Proceedings, 2018(1).
- Witzel, M. (2003). Vedas and Upanisads. In G. Flood (Ed.), *The Blackwell Companion to Hinduism* (pp. 68-101). Blackwell Publishing. (Daje pregled vedske književnosti i konteksta prijenosa).
- Wolpaw, J. R., & Wolpaw, E. W. (ur.). (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press.
- Wooldridge, M. (2009). *An Introduction to MultiAgent Systems* (2nd ed.). John Wiley & Sons.
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, Z., ... & Wang, C. (2023). Autogen: Enabling next-gen llm applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*.
- Wu, S., He, K., Wang, G., Liu, X., Chen, Z., ... & Tung, F. (2024). Exploring the clinical applications of large language models. *Intelligent Medicine*, 4(2), 100-109. <https://doi.org/10.1016/j.imed.2024.02.001>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. arXiv preprint arXiv:1609.08144.
- Wu, Y., Wang, S., Wang, T., Wang, T., Wang, H., & Yang, Q. (2021). *Collaborative Writing with GPT-3: A Preliminary Study. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Article No.: 316.
- Xiao, L., & Kumar, V. (2021). Robotics for customer service: A useful complement or an ultimate substitute? *Journal of Service Research*, 24(1), 9-29. <https://doi.org/10.1177/1094670520939886>
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing Reasoning and Acting in Language Models. *Proceedings of the International Conference on Learning Representations (ICLR) 2023*.
- You, Y., Li, J., Reddi, S., Hseu, J., Kumar, S., Bhojanapalli, S., ... & Demmel, J. (2020). *Large Batch Optimization for Deep Learning: Training BERT in 76 minutes*. *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1-18.
- Yudkowsky, E. (2008). *Artificial Intelligence as a Positive and Negative Factor in Global Risk*. U *Global Catastrophic Risks* (str. 308–345). Oxford University Press.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27.

- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). *A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), 39–58.
- Zhang, L., Wang, Y., Chen, M., & Xiao, J. (2023). Large Language Models for Mental Health: A Systematic Review. arXiv preprint arXiv:2312.04123. <https://arxiv.org/abs/2312.04123>
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). *Deep Learning Based Recommender System: A Survey and New Perspectives*. ACM Computing Surveys, 52(1), 1–38.
- Zhang, Y., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., ... & Yang, Z. (2021). *OPT: Open Pre-trained Transformer Language Models*. arXiv preprint arXiv:2205.01068.
- Zhang, Y., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., ... & Yang, Z. (2021). *OPT: Open Pre-trained Transformer Language Models*. arXiv preprint arXiv:2205.01068.
- Zhang, Y., Sun, S., Galley, M., Chen, Y. C., Brockett, C., Gao, X., ... & Dolan, B. (2020). *Dialogpt: Large-scale generative pre-training for conversational response generation*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 270–278.
- Zhang, Y., Sun, S., Galley, M., Chen, Y. C., Brockett, C., Gao, X., ... & Dolan, B. (2020). *DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 270–278.
- Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., & Dolan, B. (2020). Dialog State Tracking: A Neural Reading Comprehension Approach. *ACL 2020*.
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2021). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *NAACL-HLT 2021*.
- Zheng, Y., Shah, D. V., Cappella, J. N., & Lariscy, R. (2018). Online Political Mobilization. In K. H. Jamieson & F. F. Jr. (Eds.), *The Oxford Handbook of the Science of Science Communication*. Oxford University Press.
- Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). *The Design and Implementation of Xiaoice, an Empathetic Social Chatbot*. *Computational Linguistics*, 46(1), 53–93.
- Zhou, W., Long, F., Zhou, Q., & Huang, X. (2022). Learning Sociolinguistic Style in Dialogue Agents. *Findings of EMNLP 2022*.
- Zou, C., Li, G., & Yang, Y. (2021). Learning to Optimize Classroom Teaching with Conversational AI. *ICML 2021*.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.