

PROMPT ENGINEERING

Pisanje upita / PROMPTA , praksa je pažljivog dizajniranja i usavršavanja ulaznih upita ili zahtjeva koji se daju programu jezičnog modela kako bi se postigle željene izlazne informacije ili odgovori.

Prompt je uvijek u obliku texta

Prompt je ono **ŠTO** želimo da AI napravi

Primjeri jezičnih modela ili LLM (Large Language models) ili text to text su

- ChatGPT
- Gemini
- Claude...

Prompt Engineering je ključan za dobivanje maksimuma iz jezičnih modela jer **kvaliteta upita izravno utječe na kvalitetu odgovora modela.**



GARBAGE IN GARBAGE OUT

Postoje i drugi LLM modeli poput onih koji su

- text to audio
- text to video
- text to music
- action transformers - text to action

Ovi modeli rade na par načina

- **Reinforcement Learning with human feedback** ili RLHF gdje se modeli treniraju pomoću nagrada i kazni zavisno o tome dali daju dobar ili loš odgovor
- Text to image ili bilo koja multimedia sada radi na **Diffusion models** - to su modeli koji kreću s praznim papirom i postepeno pune papir/canvas baš kao Midjourney

Najbitnije je da jednom izgrađeni modeli mogu lagano u svoju bazu implementirati nova znanja što se zove **ZERO SHOT LEARNING** te se ne moraju ponovno učiti.



Student studying online , Focused, Self-paced, Digital Literacy, Home Workspace, Smartphone, Close-up, Evening, Relaxed, Digital

ŠTO JE BITNO ZA DIZAJN

KVALITETNOG PROMPTA

- **Jasnoća i specifičnost:** Upiti bi trebali biti jasni i specifični u vezi s onim što želite od modela.

Nejasni ili dvosmisleni upiti mogu dovesti do manje relevantnih odgovora.

- **Kontekst:** Pružanje konteksta unutar upita/prompta može pomoći modelu da bolje razumije zadatak koji mu je postavljen.

Kontekstualne informacije mogu uključivati pozadinske informacije, relevantne detalje ili ograničenja.

Primjer.

Ja sam marketing manager s 15 godina iskustva iz Osijeka...

- **Upute:** Jasno uputite modelu kako pristupiti zadatku. Možete specificirati format odgovora, zamoliti model da razmišlja korak po korak, zatražiti sažetak itd.

Primjer.

Objasni mi kako radi Google tražilica korak po korak

- **Ton i stil:** Također možete specificirati ton ili stil u kojem želite da odgovor bude. Ton ili stil može biti formalan, neformalni, veseo ili tužan...

Ton ili stil pomaže prilagodbi rezultata našim potrebama. Ton daje osjećaj.

Primjer.

Napiši mi mail o nadolazećem webinaru koristeći [prijateljski, motivirajući, pozitivan...] ton

- **Iterativni proces:** Kreiranje upita/prompta često uključuje iterativan pristup.

Ako niste zadovoljni početnim odgovorom, možete usavršiti i preformulirati upit kako biste dobili željeni rezultat.

Primjer.

Nakon odgovora napišite "Kako možeš poboljšati ovaj odgovor?"

Loše napisani i organizirani promptovi neće dati kvalitetan odgovor i mogu biti neshvatljivi.

U LLM modelima poput ChatGPT (ukoliko pričate o različitim temama u ISTOM chatu) program se zna "zablesati" i krenuti davati nepovezane odgovore.

To se većinom dogodi nakon nekih 20 upita u istom chatu. Najbolje je započeti novi chat s novom temom i držati se iste.

Kreiranje kvalitetnog upita je ključna vještina prilikom rada s jezičnim modelima jer omogućava korisnicima da učinkovito iskoriste mogućnosti ovih modela i dobiju smislene i točne informacije ili odgovore.

Prompt može biti

INSTRUKCIJA - napiši mi 3 stiha o ljubavi

PITANJE - što kupiti curi za rođendan?

PODATCI - Marija iz Vukovara ima 25 godina, Marko iz Osijeka ima 27 godina. Napiši ljubavnu priču od 300 riječi o njima.. i drugo

Promptovi mogu uključivati više elemenata

INSTRUKCIJA + PITANJE

- Instrukcija: *Napiši mi 3 stiha o ljubavi.*
- Pitanje: *Kako bih mogao upotrijebiti te stihove da izrazim svoje osjećaje prema partneru?*

Ovaj prompt kombinira zahtjev za kreativnim sadržajem (stihovi o ljubavi) s praktičnim pitanjem o njegovoj upotrebi. Odgovor bi uključivao i stihove i savjete o tome kako ih koristiti.



PITANJE + INSTRUKCIJA

- Pitanje: *Što kupiti curi za rođendan?*
- Instrukcija: *Daj mi tri originalne ideje za poklone i kratko objasni zašto bi svaka od njih bila dobar izbor.*

Ovdje se traže konkretni prijedlozi za poklon, uz obrazloženje za svaki prijedlog, što pomaže u donošenju odluke.

PODATCI + INSTRUKCIJA

- Podatci: *Marija iz Vukovara ima 25 godina, Marko iz Osijeka ima 27 godina.*
- Instrukcija: *Napiši ljubavnu priču od 300 riječi o njima koja uključuje kako su se upoznali na koncertu i kako zajedno putuju Europom.*

Ovaj prompt zahtijeva stvaranje priče na temelju zadanih podataka. Sadržaj bi bio kreativan, ali i prilagođen specifičnim informacijama o likovima.

INSTRUKCIJA + PITANJE

- Instrukcija: *Želim napraviti origami.*
- Pitanje: *Odakle krenuti i koje su osnovne tehnike koje trebam znati?*



RAZNI LLM MODELI

Postoje razni modeli za raznu upotrebu, baš kao i mobiteli. Ne morate ih razumijeti u potpunosti da bi ih znali koristiti ali svaki ima neke svoje specifične prednosti.

Postoje razni LLM modeli poput:

- **OPĆENITI:** Model treniran na velikom broju texta poput ChatGPT koji može pisati text, odgovarati na pitanja, prevoditi text itd..
- **SPECIJALIZIRANI:** Model koji je treniran na SPECIFIČNIM domenama. Npr. Model za medicinske potrebe koji može postaviti dijagnozu bolesti.
- **KONTROLIRANI:** Model koji je posebno treniran za stvaranje posebnih simbola texta npr. model koji piše pjesme u stilu i tonu Elvisa.

Ovi modeli se većinom baziraju na kreativnost



JEDNOSTAVNI I DETALJNI PROMPTOVI

Promptovi mogu biti i vrlo jednostavni ali onda očekujete i jednostavan (općeniti) odgovor.

Primjer.

Velika je razlika ako napišete u ChatGPT ova 3 prompt

- **Želim smršaviti**
- **Želim smršaviti 3 kg u mjesec dana** (dodana mjerna jedinica i vremenski period)
- **Imam 39 godina, težak sam 110 kg te visok 2 metra. Želim smršaviti 3 kg u mjesec dana. Dnevno imam 30 minuta za trening. Ponašaj se kao vrhunski fitness trener i napravi mi trening i plan ishrane za mjesec dana kako bi postigao svoj željeni cilj. Stavi trening i ishranu u tablicu.**

(dodan kontekst, mjerna jedinica, vremenski period, uloga, format)

CHAIN OF THOUGHTS PROMPTOVI

Chain of Thoughts promptovi potiču poboljšanje sposobnosti modela pri zaključivanju.

To se radi rastavljanjem složenih zadataka u niz manjih koraka kojima se lakše upravlja.

Ovakav pristup omogućuje LLM-u da se usredotoči na pojedinačne korake i nauči kako o njima učinkovitije razmišljati.

Primjer.

Tko je i gdje osvojio NBA naslov 1998?
Koristi ovaj format:
Q: <ponovi pitanje>
A: Idemo korak po korak. <objasni> i zato <finalni odgovor>



You

Tko je i gdje osvojio NBA naslov 1998?

Koristi ovaj format:

Q: <ponovi pitanje>

A: Idemo korak po korak. <objasni> i zato <finalni odgovor>



ChatGPT

Q: Tko je i gdje osvojio NBA naslov 1998?

A: Idemo korak po korak. U sezoni NBA 1997-1998, Chicago Bulls su osvojili NBA naslov. Finale NBA lige odigrano je protiv Utah Jazza. Prva dva susreta odigrana su u Chicagu, a slijedeća tri u Salt Lake Cityju. Nakon izjednačene borbe, Chicago Bulls su pobijedili u šestoj utakmici serije, osvojivši naslov prvaka. Michael Jordan bio je ključni igrač Bulls-a u tom finalu, a završna serija označila je njegovu posljednju utakmicu u dresu Chicaga prije umirovljenja.



TOKENI I RIJEČI

U GENERATIVE AI često ćete čuti spominjanje riječi **tokeni**.

Što su tokeni.

Token možemo zamisliti kao malenu jedinicu koju veliki jezični model lako može razumjeti.

U kontekstu modela poput ChatGPT, "token" se odnosi na najmanju jedinicu teksta koju model obrađuje. U jezičnim modelima poput GPT-3 i GPT-4, tokeni su obično riječi, dijelovi riječi, ili čak pojedinačni znakovi (kao što su slova ili interpunkcijski znakovi), ovisno o tome kako je model treniran.

Tolken može biti bilo koji dostupan simbol poput , . ! ()...

Primjer.

Volim igrati košarku - **3 tokena**

Jedna riječ može biti sačinjena od više tokena. Neke riječi su jedan token, a neke više.

Riječ
EVERYDAY
JOYFULL

Tolken
Every, day
Joy, full

Različiti LLM modeli imaju različite mehanizme kojima razdvajaju input na tokene. Ovaj korak poznat je kao **tokenizacija**. Često, metoda tokenizacije zaista mijenja rezultate modela.

Model koji se koristi postavlja ograničenje tokena za zahtjeve.

Besplatna verzija ChatGPT-3.5 maksimalno može koristiti **4097** tokena **koji se dijele između upita i odgovora**.

Ako vaš prompt koristi 3000 tokena, možete generirati najviše 1097 tokena odgovora.

ChatGPT - 3.5
Ima **4096** tokena

ChatGPT - 4 (plaćena verzija)
Dopušta **128,000** tokena

Tokeni su temeljni elementi teksta u procesu obrade prirodnog jezika, omogućujući modelima poput ChatGPT-a da efikasno komuniciraju i odgovaraju na upite.



KAKO LLM RADE

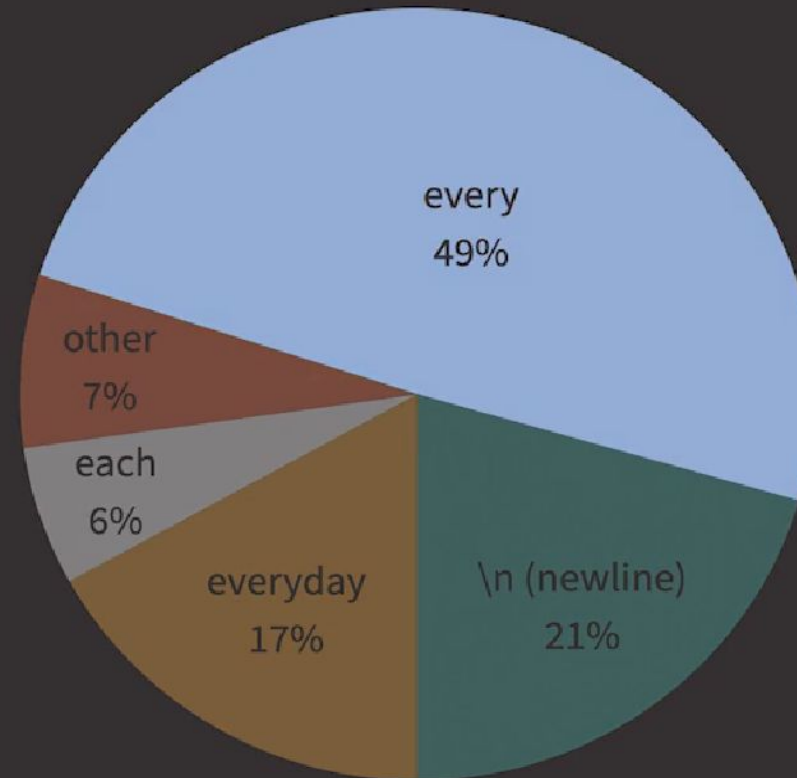
*Danas je
divan _____*

dovrši rečenicu

KAKO LLM RADE

Prompt:

“I try to learn something new”



Umjetna inteligencija pokušava “samo” odgonetnuti koja bi bila sljedeća riječ ili simbol. Za tu vrstu podatka potrebno je ogromna količina sličnih podataka (texta) kako bi se dobila ova vjerojatnost iduće riječi.

PRIMJER PROMTOVA

[CILJ], [KONTEKSTS], [AKCIJA]

Prvo trebamo znati koji nam je cilj i biti što detaljniji

Primjer.

Želim imati 1000 pratitelja na Facebooku do 10. travnja 2025.

Napravio sam web stranicu koja nudi online edukacije na temu **umjetne inteligencije i marketing**. Želim da na mojoj Facebook stranici ljudi dolaze radi informiranja o najnovijim vijestima iz **AI i marketing svijeta**. Za sada nemam pratitelja na facebook stranici ali **imam ažurnu web stranicu i ažuran blog** na istoj.

Kreiraj mi content strategiju, ciljnu publiku i ciljeve te razne preporuke kako mogu doći do mog cilja.

BITNO ZA ZNATI

- **BUDITE ŠTO PRECIZNIJI / DATALJNIJI**
- **PRIČAJTE KAO DA PRIČATE S PRIJATELJEM.**

Svi LLM modeli su bazirani na prirodnom jeziku i nemojte se zavaravati da morate zvučati pametno.

- **POSTAVLJAJTE POD PITANJA**

Uvijek možete pitati za dodatno pojašnjenje odgovora te ukoliko niste zadovoljni možete (si) postaviti ova dodatna pitanja.

- Analiziraj dodatno odgovor
- Obrazloži dodatno odgovor
- Proširi dodatno odgovor

- **NEMOJTE KORISTITI “KAKO SI” , “POZDRAV”, “MOLIM TE”...**

To Vam sve troši tolkene nepotrebno

VIŠE O PROMPTOVIMA SAZNAJTE NA
OVOM LINKU

[OpenAI Prompt](#)