

DALL E 2 - OpenAI

Seminarski rad u okviru kursa
Tehničko i naučno pisanje
Matematički fakultet

Ognjen Marković

ognjen.mark03@gmail.com

Matija Milovanović

matija.milovanovic55@gmail.com

Bogdan Milovanović

bogdan.milovanovic1106@gmail.com

Bojana Radovanović

bojanaradovanovic080@gmail.com

11. novembar 2022.

Sažetak

DALL-E I DALL-E 2 su modeli mašinskog učenja koje je razvio OpenAI za generisanje digitalnih slika pomocu tekstualnih opisa generisanih metodama obrade prirodnih jezika (eng. Natural language processing - NLP). DALL-E je prvi put spomenut na blogu OpenAI-a u januaru 2021. i tada je koristio verziju GPT-3 modifikovanu za generisanje slika. U aprilu 2022. OpenAI je najavio DALL-E 2, naslednika dizajniranog da generiše realističnije slike u višim rezolucijama koje mogu kombinovati koncepte, atribute i stilove.

Sadržaj

1	Uvod	2
2	Veštačka inteligencija	2
3	DALL-E 2	3
3.1	Tehnologija	3
3.2	Mogućnosti	4
3.3	Etički problemi	6
3.4	Tehnička ograničenja	6
4	Zaključak	7
	Literatura	7

1 Uvod

OpenAI je istraživačka laboratorija za veštačku inteligenciju ((eng. Artificial Intelligence - AI)) koju čine profitna korporacija OpenAI LP i njena matična kompanija, neprofitna OpenAI Inc. Kompanija, koja se smatra konkurentom DeepMind-¹, sprovodi istraživanja u oblasti veštačke inteligencije sa navedenim ciljem promovisanja i razvoja prijateljske veštačke inteligencije na način koji koristi čovečanstvu u celini. Organizaciju su osnovali u San Francisku krajem 2015. godine Ilon Mask, Sam Altman i drugi, koji su zajednički obećali milijardu dolara. Mask je podneo ostavku na mestu u odboru u februaru 2018. godine, ali je ostao donator. U 2019. godini, OpenAI LP je od Microsoft-a dobio investiciju od milijardu dolara. Neki naučnici, poput Stivena Hokinga i Stjuarta Rasela, izrazili su zabrinutost da ako napredna veštačka inteligencija jednog dana dobije sposobnost da se redizajnira sve većom brzinom, nezaustavljava „eksplozija inteligencije“ može dovesti do izumiranja ljudi. Mask karakteriše veštačku inteligenciju kao „najveću egzistencijalnu pretnju čovečanstva“. Osnivači OpenAI su je strukturisali kao neprofitnu organizaciju kako bi mogli da fokusiraju svoje istraživanje na stvaranjepozitivnog dugoročnog uticaja na ljude[1].

2 Veštačka inteligencija

Veštačka inteligencija jedna je od retkih oblasti nauke o kojoj skoro svi – i eksperti i oni koji to nisu – imaju neki stav. Neko smatra da ona donosi velike koristi, neko smatra da od nje prete opasnosti, a neko veruje i u jedno i u drugo. Neobično je onda da, s druge strane, ne postoji opšta saglasnost o tome šta je uopšte veštačka inteligencija i čime se ona bavi. Pod inteligencijom se obično podrazumeva sposobnost usvajanja, pamćenja i obrade određenih znanja. Iako postoji i shvatnje po kojem je centralni cilj veštačke inteligencije oponašanje ljudske inteligencije, većina podoblasti veštačke inteligencije ima drugačiji cilj. Obično je to rešavanje problema u kojima se javlja kombinatorna eksplozija, tj. u kojima je broj mogućnosti toliko veliki da se ne može sistematicno ispitati u razumnom vremenu. Najveći deo zadataka veštačke inteligencije može se opisati u terminima algoritmike, pretrage, deduktivnog i induktivnog zaključivanja, i drugih preciznih matematičkih pojmoveva. Tek mali deo istraživača bavi se metodama koje pretenduju da dostignu opšte rasuđivanje u stilu čoveka. Veštačka inteligencija sastoји se od više podoblasti, naizgled slabo povezanih svojim sadržajem. Pitanje je onda šta je to zajedničko za njih, sem to što se uglavnom sve bave problemima u kojima se javlja mnoštvo mogućnosti koje se ne mogu ispitati sistematičnom pretragom. Termin "veštačka inteligencija" je ranije korišćen za opisivanje mašina koje oponašaju i prikazuju ljudske kognitivne veštine koje su povezane sa ljudskim umom, kao što su učenje i rešavanje problema. Ovu definiciju su od tada odbacili glavni istraživači AI koji sada opisuju AI u smislu racionalnosti i racionalnog delovanja, što ne ograničava način na koji se inteligencija može artikulisati.

¹DeepMind je kompanija koja uz pomoć veštačke inteligencije kreira programe koji koriste duboke neuronske mreže da bi naučili sebe kako da igraju različite igre kao što su Go i šah. Osnovana je 2010. godine, a već 2014. godine je kupljena od strane Google-a.

Primena veštačke inteligencije je jako široka i ona uključuje napredne web pretraživače (npr. Google), sisteme za preporuke (koje koriste YouTube, Amazon i Netflix), razumevanje ljudskog govora (kao što su Siri i Alexa), automobile koji se samostalno voze (npr. Tesla), automatizovano donošenje odluka i takmičenje na najvišem nivou u sistemima strateških igara (kao što su šah i Go). Kako mašine postaju sve sposobnije, zadaci za koje se smatra da zahtevaju „inteligenciju“ često se uklanjuju iz definicije veštačke inteligencije.

Takođe prisutna je još jedna oblast veštačke inteligencije koje će nama biti jako interesantna, a to je NLP. Ona se bavi interakcijom između računara i ljudskog jezika, posebno kako programirati računare da obrađuju i analiziraju velike količine teksta na prirodnom jeziku. Cilj NLP-a je da kompjuter razume sadržaj dokumenata, uključujući kontekstualne nijanse jezika u njima. Tehnologija tada može precizno da izdvoji informacije sadržane u dokumentima, kao i da kategorizuje i organizuje same dokumente^[2].

3 DALL-E 2

3.1 Tehnologija

Model generativnog unapred obučenog transformatora (eng. Generative Pretrained Transformer - GPT) je prvobitno razvio OpenAI 2018. godine, koristec arhitekturu transformatora. Prva iteracija, GPT, je unapređena da bi proizvela GPT-2 2019. godine; 2020. ponovo je unapređena da bi proizvela GPT-3, sa 175 milijardi parametara. DALL-E-ov model je multimodalna implementacija GPT-3 sa 12 milijardi parametara koji „zamenjuje tekst za piksele“, obučen na parovima tekst-slika sa Interneta. DALL-E 2 koristi 3,5 milijardi parametara, što je manji broj od svog prethodnika. Naziv softvera je sastavni deo imena animiranog robota Pikar lika

Tabela 1: U tabeli 1 je prikazan hronološki razvitak DALL-E 2.

2018	2019	2020	2021	2022
GPT	GPT-2	GPT-3	DALL-E	DALL-E 2

VALL-E i španskog nadrealističkog umetnika Salvadoru Dalija. DALL-E 2 je razvijen i objavljen javnosti u saradnji sa CLIP-om (eng. Contrastive Language-Image Pre-Training - CLIP). Njegova uloga je da „razume i rangira“ sliku koja je izlaz iz DALL-E predviđanjem koji je naslov sa liste od 32.768 naslova nasumično odabranih iz skupa podataka (od kojih je jedan bio tačan odgovor) najprikladniji za sliku. Ovaj model se koristi za filtriranje već početne liste slika koje generiše DALL-E da bi se odabrali najprikladniji rezultati. Kad bismo pojednostavili stvari, ovaj model funkcioniše tako što počinje sa slikom koja se sastoji od nasumičnih piksela („šum“) i postepeno „uklanja šum“ slike. Tokom procesa uklanjanja šuma, on se približava slici koja odgovara početnom izvornom tekstu koji je dat na ulazu.

3.2 Mogućnosti

DALL-E 2 može da:

- generiše slike u više stilova, uključujući fotorealistične slike i emotikone
- da manipuliše i preuređuje objekte na svojim slikama
- ispravno postavi elemente dizajna u nove kompozicije bez eksplisitnih instrukcija

Thom Dunn koji je pisao za BoingBoing je primetio da, na primer, "kada se zamoli da nacrti daikon rotkvu kako duva nos, pijucka kafu ili se vozi monociklom, DALL-E često crta maramicu, ruke i stopala na uverljivim mestima". DALL-E je pokazao sposobnost da „popuni prazna mesta“ kako bi zaključio odgovarajuće detalje bez specifičnih napomena kao što je dodavanje božićnih slika uputstvima koja se obično povezuju sa proslavom, i odgovarajuće postavljene senke slikama koje ih ne pominju. Šta više, DALL-E pokazuje široko razumevanje vizuelnih dizajnerskih trendova[3].

DALL-E je u stanju da proizvede slike za širok spektar proizvoljnih opisa sa različitih gledišta uz minimalan procenat grešaka. Mark Ridl, vanredni professor na Tehničkoj školi za interaktivno računarstvo u Džordžiji, otkrio je da DALL-E može da kombinuje koncepte (opisan kao ključni element ljudske kreativnosti) i ima sposobnost vizuelnog rasudivanja dovoljnu da reši Rejvenove matrice (vizuelni testovi koji se često primenjuju ljudima za merenje inteligencije).

DALL-E 2 može da kreira veoma realistične slike kao npr. prikazan portret basketa (slika 1.).



Slika 1: Pas

Ali trik u svemu ovome je što ovaj pas ne postoji.. Ova slika psa nije nastala tako što je osoba slikala svog četvoronočnog prijatelja, već je slika koju je napravilo DALL-E 2. Takođe, u stanju je da kreira veoma realistične slike ljudi npr. prikaz realne fotografije mlade žene plavih očiju i plave kose (slika 2.) ,kao i objekte koji u stvarnosti ne postoje.



Slika 2: Žena

Neki od zanimljivijih nemogućih slika su: "Hamburger u obliku Rubikove kocke, profesionalno fotografisanje hrane" (slika 3.), "Plišani medvedi rade na novom istraživanju veštačke inteligencije na Mesecu 1980-ih" (slika 4.)...



Slika 3: Rubikova kocka kao hamburger



Slika 4: Medvedi u svemiru

3.3 Etički problemi

Oslanjanje DALL-E 2 na javne skupove podataka utiče na njegove rezultate i dovodi do algoritamske pristrasnosti u nekim slučajevima, kao što je generisanje većeg broja muškaraca nego žena za zahteve koji ne pominju pol. Podaci o obuci DALL-E 2 su filtrirani kako bi se uklonile nasilne i seksualne slike, ali je utvrđeno da to povećava pristrasnost u nekim slučajevima, kao što je smanjenje učestalosti generisanja žena. OpenAI prepostavlja da je to možda zato što je veća verovatnoća da će žene biti seksualizovane u podacima o obuci, što je uzrokovalo da filter utiče na rezultate. U septembru 2022., OpenAI je potvrdio za The Verge da DALL-E nevidljivo ubacuje fraze u korisničke upite kako bi se pozabavio pristrasnošću u rezultatima; na primer, „crnac“ i „azijtska žena“ se ubacuju u upite koji ne navode pol ili rasu.

Zabrinutost u vezi sa DALL-E 2 i sličnim modelima za generisanje slika je da bi se oni mogli koristiti za propagiranje dubokih fejkova² i drugih oblika dezinformacija. Kao pokušaj da se ovo ublaži, softver odbija upite koji uključuju javne ličnosti i rezultate koji sadrže ljudska lica. Upiti koji sadrže potencijalno nepoželjan sadržaj se blokiraju, a otpremljene slike se analiziraju kako bi se otkrio uvredljiv materijal. Nedostatak filtriranja zasnovanog na upitu je to što ga je lako zaobići upotrebot alternativnih fraza koje rezultiraju sličnim izlazom. Na primer, reč „krv“ je filtrirana, ali „kečap“ i „crvena tečnost“ nisu, pa korisnici na taj način mogu dobiti sličnu sliku izbegavši filtriranje[4].

Još jedna zabrinutost u vezi sa DALLE-2 i sličnim modelima je da bi mogli da izazovu tehnološku nezaposlenost za umetnike, fotografе i grafičke dizajnere zbog svoje tačnosti i popularnosti.

3.4 Tehnička ograničenja

Razumevanje jezika DALL-E 2 ima ograničenja. Ponekad nije u stanju da razlikuje žutu knjigu i crvenu vazu od crvene knjige i žute vase ili pandu koja pravi latte art od late umetnosti pande. Generiše slike astronauta koji jaše konja kada mu se prikaže upit konja koji jaše astronauta. Tako de ne uspeva da generiše ispravne slike u različitim okolnostima. Zahtevanje više od 3 objekta, negacija, brojevi i povezane rečenice mogu dovesti do grešaka i karakteristike objekta mogu se pojaviti na pogrešnom objektu. Dodatna ograničenja uključuju rukovanje tekstom – koji, čak i sa čitljivim slovima, gotovo uvek ispadne kao besmislica u obliku snova– i njegov ograničeni kapacitet da se bavi naučnim informacijama, kao što su astronomija ili medicinske slike.

²Dezinformacije putem slika.

4 Zaključak

Literatura

- [1] Allyn, Bobby. *Surreal or too real? Breathtaking AI tool DALL-E takes its images to a bigger stage.* NPR. Retrieved 20 July 2022.
- [2] OpenAI *DALL-E 2 Preview - Risks and Limitations*, 19 June 2022
- [3] Macaulay, Thomas *Say hello to OpenAI's DALL-E, a GPT-3-powered bot that creates weird images from text.* TheNextWeb, 28 January 2021.
- [4] Sahar Mor, Stripe *"How DALL-E 2 could solve major computer vision challenges".*
- [5] Walsh, Bryan *"A new AI model draws images from text"*