

AI koji generiše sliku na osnovu teksta

Budimir Nikola 178/2021 Trajković Miljan 354/2022

Cvejić Miloš 346/2021 Bajić Bogdan 122/2021

5. decembar 2022.

Sažetak

Ovaj rad je napisan sa ciljem da vas upozna sa jednom aktuelnom primenom veštačke inteligencije. Ukratko će vam biti objašnjeno kako se ovakav sistem razvija, kako funkcioniše i kako ga i vi možete koristiti, a zatim vas će i sprovesti kroz probleme koje donosi sa sobom.

Sadržaj

1	Uvod	2
2	Kako funkcionišu AI generatori slika	4
2.1	Pregled arhitekture DALL-E 2 softvera	5
3	Treniranje AI generatora slika	7
4	Kako bilo ko može da koristi AI za generisanje slika	9
4.1	DALL-E 2	9
4.2	Midjourney	10
4.3	Stable Diffusion	10
5	Etički problemi sa veštačko inteligentnim generatorima slika	11
5.1	Problem autorskih prava	12
6	Zaključak	13

1 Uvod

Veštačka inteligencija

Veštačka inteligencija, iliti AI (skraćeno od eng. artificial intelligence) predstavlja podoblast računarstva koja omogućava računarima da realizuju funkcije koje su karakteristika ljudskog razmišljanja, odnosno da što vernije imitiraju rad ljudskog mozga. Veštačka inteligencija je počela da se razvija tokom druge polovine dvadesetog veka, a tokom proteklih godina taj razvoj je počeo da se sve više i više ubrzava. Veštačka inteligencija se, iako je još uvek tek u svojim početnim fazama, koristi u raznim oblastima za rešavanje mnogih problema. [1]

Neka od čestih mesta na kojima se upotrebljava veštačka inteligencija su obrada podataka, prevođenje jezika, prepoznavanje slika, razvoj softvera, elektronska trgovina (e-commerce), video igre, medicina, napredni internet pretraživači, sistemi preporuke (poput onih koje koriste YouTube, Amazon, Netflix i slični), robotika, proizvodnja bespilotnih letelica i autonomnih automobila (poput onih koje proizvodi Tesla), generisanje slika, kao i još mnogo drugih.

Mašinsko učenje predstavlja podoblast veštačke inteligencije. Ono omogućava da računar “nauči” zadati šablon koji dobija kroz milione podataka i upotrebi ga kako bi mogao da oponaša taj isti tip ponašanja.

Pristupi mašinskom učenju se generalno dele u tri različite kategorije (koje odgovaraju paradigmama učenja), u zavisnosti od prirode signala ili povratne informacije dostupne sistemu učenja. To su: učenje pod nadzorom (eng. supervised learning), polunadgledano učenje (eng. semi-supervised learning), učenje bez nadzora (unsupervised learning) i učenje potkrepljenjem (reinforcement learning).

Učenje pod nadzorom je paradigma mašinskog učenja kod koje postoji nadgledač koji predstavlja učitelja. Kod ovog tipa učenja, obučavamo računar koristeći podatke koji su ispravno označeni, nakon čega mu dajemo pristup novim podacima, koje on treba da analizira, prepozna i da zna šta treba da radi sa njima. Učenje bez nadzora je paradigma mašinskog učenja kod koje se računar obučava koristeći informacije koje nisu označene (za razliku od učenja pod nadzorom). Zadatak računara je da grupiše zadate podatke prema njihovim sličnostima, bez ikakve prethodne obuke. Učenje potkrepljenjem predstavlja paradigmu mašinskog učenja koja podseća na treniranje kućnog ljubimca. Radi se o nagrađivanju računara kada je radnja koju je učinio

adekvatna. Na prvi pogled to podseća na učenje pod nadzorom, ali za razliku od tog tipa, kod učenja potkrepljenjem računar nema podatke koje koristi za buku, već uči iz sopstvenog iskustva. [2]

Šta je AI generator slika

AI generator slika predstavlja program koji generiše sliku na osnovu zahteva zadatog u obliku teksta koristeći veštačku inteligenciju. To funkcioniše tako što generator slika ima pristup bazi podataka u kojoj se nalazi veliki broj različitih slika, na osnovu kojih on uči da generiše nove slike. Kada generator slika dobije opis slike koju treba da generiše, on u bazi podataka proverava slike sa sličnim opisom i na osnovu njih generiše sliku koja podseća na njih.

Slike generisane korišćenjem veštačke inteligencije često se koriste za pravljenje reklama, postera, imitaciju poznatih dela, kao i za dobijanje inspiracije, ili prosto radi zabave.

Neki od poznatih programa za generisanje slika na osnovu zadatog teksta su:

- Jasper Art
- Canva AI Art
- Starry AI
- Nightcafe
- Dall-E
- Pixray
- Deep Dream Generator
- Deep AI
- ArtBreeder
- Hypotenuse AI
- Craiyon (DALL-E mini)
- i drugi.

[3]

2 Kako funkcionišu AI generatori slika

Preteče današnjih generatora

Mnogi modeli veštačke inteligencije pa tako i generatori slika na osnovu teksta poput **DALL-E 2** i drugih, zasnovani su na prethodnim istraživanjima i prethodnim modelima iz ove oblasti. Generisanje slika razvijalo se linearno i to prvobitno koriscenjem mreža koje koriste metode suparničkog učenja odnosno “Generative adversarial network” (GANs).

GANs su modeli mašinskog učenja obično sačinjeni od dve mreže, *diskriminatora* i *generatora*, koje međusobnim intereagovanjem treniraju jedna drugu. Naime, ako uzmemo na primer generisanje slika, zadatak diskriminatora je da što sigurnije i preciznije odredi da li je njemu zadata slika, prava fotografija iz neke baze podataka ili je to generisana slika. Sa druge strane, zadatak generatora je da prevari diskriminatora, odnosno da što bolje nauči šta sliku čini “pravom” i tako napravi sliku za koju se ne može utvrditi da li je ona prava ili ne. U tom slučaju interakcija ove dve mreže i njihovo treniranje se može posmatrati kao minimaks igra u kojoj se diskriminator trenira da maksimalno uveća verovatnoću tačnog određivanja klase slika, dok se generator trenira da maksimalno umanji verovatnoću da slike koje je generisao budu klasifikovane od strane diskriminatora kao generisane slike.[4, 5, 6]

Ključna razlika između ovako opisanih mreža i najnovijih AI generatora slika je ta što se klasični GAN modeli obično razvijaju za neku konkretnu kategoriju fotografija(npr. popularni generator ljudskih lica)[7] dok na drugoj strani generatori poput DALL-E 2, Stable Diffusion, Imagen-a su mnogo opštije namene i možemo im zadati tekstualni opis slike koju očekujemo od generatora. To zadavanje opisa se realizuje proširivanjem već pomenuutih mreža tako što se generatoru prosleđuje uslov po kom će on generisati sliku.[8, 9, 11] Formiranje ovog uslova je zapravo netrivialna stvar jer najbolje realizacije generatora nisu one u kojima se kao uslov prosleđuje samo opis slike što je i pokazano u istraživanju **OpenAI** tima.[8]

Arhitekture najnovijih generatora zasnovane su na jednoj od novijih tehnika nazvanoj difuzija(eng. diffusion) koja je inspirisana oblašću iz fizike - neravnotežna termodinamika. Naime, ako sipamo kap farbe u vodu, zakoni fizike nalažu da će ta kap da *difunduje*¹ sve dok ne dođe u ravnotežu. Ovaj proces u prirodi je ireverzibilan. U mašinskom učenju, modeli difuzije nasto-

¹Difundovanje - proces spontanog raspršivanja materije i energije, prelazak iz zone više u zonu niže energije ili koncentracije

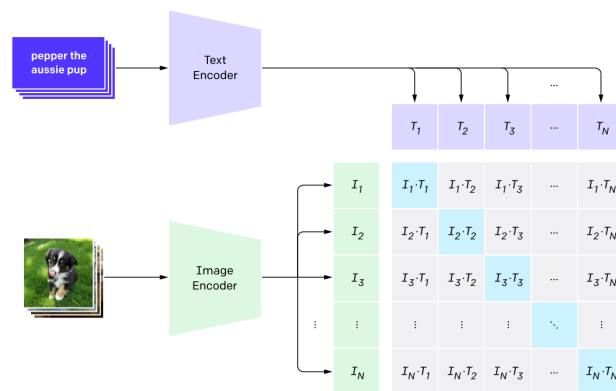
je da reše upravo ovaj problem, odnosno da nađu način kojim mogu doći do početne kapi farbe.[11]

2.1 Pregled arhitekture DALL-E 2 softvera

Na primeru konkretnog generatora DALL-E 2 možemo videti prethodno opisane mreže i koju ulogu one imaju u jednom od najpoznatijih softvera za generisanje slika.

Za pocetak posmatraćemo takozvanu CLIP mrežu (skraćeno od Contrastive Language–Image Pre-training), koja je izuzetno bitna za treniranje osnovnih komponenti ovog generatora. Njen osnovni zadatak je da zadatu sliku poveže sa odgovarajućim unetim opisom. Primetimo da se na neki način generisanje slika može smatrati suprotnim procesom od pomenutog, a razlog zašto je ova mreža od ključnog značaja prilikom generisanja ćemo videti u nastavku teksta. Na ulazu CLIP mreže nalaze se dva kodera koji kodiraju zadatu sliku i njen opis u odgovarajući *imbeding*², zasebice. Prilikom treniranja, CLIP nastoji da pomenute kodere nauči da imbeding slike i imbeding njenog opisa budu što sličniji jedan drugom.[10, 8]

1. Contrastive pre-training



Slika 1: Slika preuzeta sa OpenAI blog stranice

²imbeding (eng. embedding) - rezultat preslikavanja skupa informacija u računarski zapis, vektor

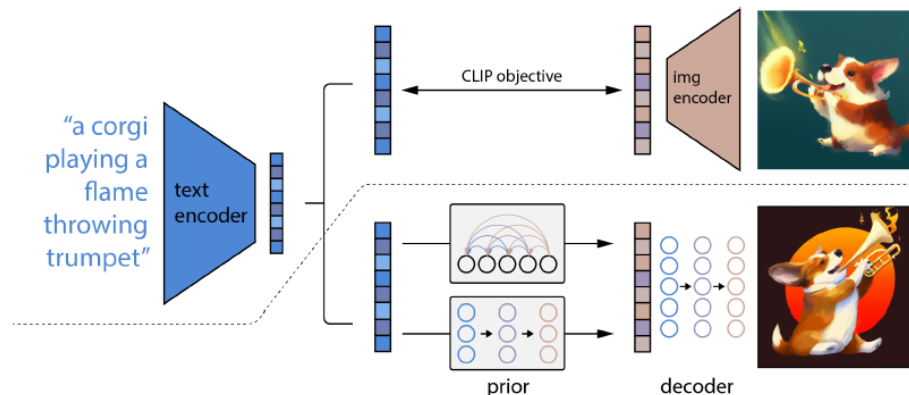
Sama arhitektura Dall-E 2 softvera se sastoji iz dva dela: Prvi deo čini mreža prior³ koji na ulazu prima tekst imbeding (formiran CLIP kodiranjem tekstualnog opisa željene slike) koji preslikava u imbeding slike koji prosleđuje drugom delu ovog generatora - dekodera. Sada dolazimo do ključnog značaja CLIP mreže. OpenAI tim se zapitao, da li je prior u ovom pristupu uopšte potreban. Iz tog razloga pokušali su da dekodera proslede samo tekstualni opis, zatim tekstualni opis i tekst imbeding, a na kraju i imbeding slike dobijen pomoću prior-a. Iz tog eksperimenta su zaključili da slike generisane imbedingom slike daju značajno bolje rezultate u odnosu na druga dva slučaja.[8]

Dekoder je generativni model realizovan pomenutom tehnikom difuzije. Naime, slike se generišu iz nasumične količine *Gausovog šuma*⁴. Zadatak generatornog dela je da pretpostavi koliko šuma je na slici i otkloni ga. Ovo je vrlo težak problem koji se uspešno rešava upravo difuzno i to na sledeći način: Otklanjanje šuma vrši se iterativno[7]. U svakoj iteraciji pretpostavlja se koliko šuma je na slici koji se otklanja, a zatim se na dobijenu sliku dodaje skoro malo manja količina Gausovog šuma od prethodnog i prelazi u sledeću iteraciju. Kako bi dekodeo znao šta uopšte treba da generiše, odnosno “šta se nalazi ispod šuma” prosleđuje mu se uslov. Kod ranijih projekata kao što su GLIDE i DALL-E[9, 8], dekodera se kao uslov direktno prosleđivao tekstualni opis željene slike. Dok se kod ovog modela zahvaljujući CLIP i prior mreži generatoru dodatno prosleđuju CLIP imbeding slike i teksta. Generisanje različitih slika i variranje u odnosu na jednu generisanu sliku se postiže tako što se dobijena slika propušta kroz CLIP koder kako bi se dobio njen imbeding slike, a onda se proces generisanja ponavlja sa novonastalim imbedingom. Na ovaj način dobija se slika koja ima iste elemente i stil prethodne slike, a različit raspored i neke trivijalne detalje koji se gube CLIP kodiranjem. Izlaz opisanog dela dekodera je slika formata 64x64.

Zato se na samom kraju ovog generatora slika rezolucije 64x64 prolazi kroz još dva ekspanzivna koraka kojim se slika dovodi na rezoluciju 1024x1024. Ovi koraci su takođe realizovani difuznim tehnikama što doprinosi tome da slika na izlazu ima više detalja i bude maksimalno fotorealistična.[8]

³prior - označava ono što je pre nečega po redu ili hronološki.

⁴Gausov šum (engl. Gaussian noise) - šum raspoređen korišćenjem Gausove normalne raspodele



Slika 2: Prikaz DALL-E 2 modela iz ptičije perspektive

3 Treniranje AI generatora slika

Kako smo već objasnili neke od modela mašinskog učenja koji se koriste u generatorima slika može samo dodati neke ključne detalje koji su bitni kod ovako velikih projekata.

Za Trening softvera kao što je generator slika neophodne su ogromne baze podataka slika koje su pažljivo birane za ovaj trening. Naravno, kreiranje ovakvih baza podataka i prikupljanje slika generalno zahteva veliki broj ljudi koji rade na ovom projektu. Jedna od najzahtevnijih baza jeste The ImageNet baza slika na kojoj je radilo preko 25 hiljada ljudi koji su skupili 14 miliona slika na kojima se može naći dvadeset dve hiljade različitih objekata. Interesantno je to da OpenAI tim nije utrošio ovakve napore prilikom treniranja Dall-E 2 generatora. [10]

Već smo pričali o CLIP mreži. Tokom njenog treniranja ona prima parove slika i njihovih opisa. Te parove mreža pronalazi na internetu. Možemo primetiti da slike dobijene Gugl pretragom, kao i slike na društvenim mrežama poput Instagrama i Tvitera gotovo uvek imaju opise i lako im se pristupa. Samim tim trošak na građenje iscrpnih baza slika potpuno je prevaziđen. [10, 8]

Treniranje dekodera svodi se na treniranje difuznog modela. Najbitnije je pomenuti jednu difuznu tehniku zvanu navođenje bez klasifikatora (eng. Classifier-free guidance). Ova tehnika podrazumeva da se generatoru u određenom broju slučajeva prosleđeni CLIP imbeding postavljaju na nulu, čime je generator forsiran da napravi sliku samo na osnovu teksta. A tokom treniranja

50% posto generacija izvršava se bez teksta, odnosno samo pomoću CLIP imbedinga.

Takođe, tokom treniranja mreža koje sliku rezolucije 64x64 proširuju na sliku rezolucije 1024x1024, slike koje dolaze na ulaz mogu biti malo zamućene korišćenjem Gausove normalne raspodele. Kao što smo rekli baš zbog ovih difuznih tehnika na izlazu dekodera dobijamo sliku rezolucije 1024x1024 koja je fotorealistična i može sadržati puno detalja.[8]

4 Kako bilo ko može da koristi AI za generisanje slika

Veštačka inteligencija koja generiše slike bila je jedna od gorućih tema u vestima ove godine zbog izuzetno brzog napretka ali i etičkih problema koje može da donese. Međutim, zašto bi, i kako, mogli da koristimo ove novotarije?

Motivacija za korišćenje:

- Može da generiše realistične, ili čak, hiperrealistične slike što se može koristiti u svrhu razvoja moderne umetnosti (npr. u filmovima za stvaranje zanimljivijih i privlačnijih scena natprirodnih pojava).
- Sposobnost konstantnog napredovanja AI-a mu dozvoljava da stalno proizvodi nove ideje, neke nikada pre viđene stvari, koje mogu biti ogromna inspiracija.

Neki od izazova:

- Bez obzira na kvalitet slika, one ipak iza sebe neće imati priču ili emocije.
- AI stvara isključivo na osnovu podataka koje poseduje što može dovesti do manjka originalnosti.
- Etički problemi koji se tiču kršenja autorskih prava, kao i raznih načina zloupotrebe veštački generisanih slika.

4.1 DALL-E 2

DALL-E 2 kompanije OpenAI jedno je od glavnih imena u ovom prostoru. Čitav proces generisanja je zapravo prilično jednostavan – unesemo opis željene stvari i dobijemo sliku toga. Ali, ima još par stavki koje treba da držimo na umu da bismo postigli savršen rezultat.

Da biste koristili DALL-E 2, sve što je potrebno jeste da se ulogujete na sajt. Dužina samog opisa tj. *prompt-a*⁵ je 400 karaktera, a sam prompt može da primi *sadržaj* i *modifikatore*. Sadržaj je npr. "*usamljeni astronaut na Marsu*", a modifikator "*misteriozno, sa puno boja, hiperrealistično*" (naravno, podrazumeva se da se reči unose na engleskom). Postoje i posebne

⁵eng. to prompt - navesti

knjige[13], kao i eksperti, koji vam mogu pomoći kako da unesete savršeni prompt. Jedan od načina da postavite atmosferu jeste da koristite reči koje predstavljaju emocije ili vremensko razdoblje. DALL-E 2 kao ulaz može da primi i sliku.

4.2 Midjourney

Midjourney je AI alatka koja funkcioniše isključivo na digitalnoj platformi Discord⁶. Ako imate nalog na Diskordu, sve što treba da uradite jeste da se prijavite preko zvaničnog sajta za Midjourney. Ostatak procesa instalacije je veoma intuitivan. Sam AI funkcioniše slično kao i prethodno pomenuti DALL-E 2 samo što se ceo proces ne odvija na sajtu već u aplikaciji. Midjourney uvek generiše 4 slične slike. Ako vam se rezultat dopao, možete da povećate rezoluciju slike, ili da tražite da generiše novu, baziranu na stilu one koje ste izabrali. Ono po čemu se razlikuje od DALL-E 2 je što se manje fokusira na realizam a više na maštovitiju i raznobojniju kreaciju.

4.3 Stable Diffusion

Svi jaki modeli za generisanje slika baziraju se na modelu dubokog učenja **Stable diffusion**. Najvažnija razlika od ostalih jeste da je on potpuno open-source⁷ što posebno doprinosi njegovom daljem razvoju, kao i čitavoj zabavi. Pored toga što je neverovatno zanimljiv za korišćenje, on takođe daje opciju da kao unos upotrebite neku ličnu sliku i manipulišete njome. Sam model možete koristiti na svom računaru preuzimanjem sa Git-a, ili na zvaničnom veb-sajtu.

DALL-E 2	openai.com/dall-e-2
Midjourney	midjourney.com
Stable Diffusion	beta.dreamstudio.ai/dream

⁶Discord(transkribovano Diskord) je platforma namenjena za komunikaciju, preventveno među ljudima koji igraju video igre.

⁷open-source - programi čiji je kod svakome dostupan za potencijalnu modifikaciju i distribuciju.



Slika 3: Poređenje ova tri AI generatora

5 Etički problemi sa veštačko inteligentnim generatorima slika

AI generatori slika mogu doprineti diskriminaciji tako što oponašaju štetne stereotipe koje je AI prikupio u kolekcijama podataka, koji sami imaju predrasude iz svakodnevnog života.

U svom članku Rune Klingenberg Hansen nudi sledeća rešenja [14]. Problem se može korigovati zabranjivanjem određenih reči ili nasumičnim dodavanjem prefiksa kao što su: žena/muškarac/trans, crna/žuta/braon osoba, kvir⁸ itd.). Ali nekada automatsko dodavanje dodatnih reči može negativno uticati umesto da reši problem. Npr. dodavanje reči žena za neke termine koji se koriste za osobe svih polova ili dodavanje reči crnac kada se spominju bande/karteke.

Još jedan problem je mogućnost veštačke inteligencije da generiše lažne informacije. To je upravo razlog zašto DALL-E [15] ne dozvoljava generisanje slika sa poznatim osobama [16], ali lažne informacije ne utiču samo na poznate ljude. Npr. ljudi se mogu svetiti svojim bivšim partnerima ili članovima svoje porodice.

Ali možda najstrašniji stvar jeste da ova tehnologija može u potpunosti uništiti naše poverenje u slike kao verodostojan dokaz. Zbog straha da su slike lažne, jer generator slika nam omogućava da kreiramo fotorealistične slike. Opasnost od ovog scenarija se može umanjiti korišćenjem tehnologija koje

⁸kvir (engl. queer) - naziv za celokupnu homoseksualnu, biseksualnu, transrodnu i interseksualnu zajednicu kao i heteroseksualne osobe koje sebe vide ili žive svoj život van heteropatrijarhalnih normi.

imaju sposobnost detektovanja lažnih slika. Takođe mogli bismo zahtevati da veštački generisane slike imaju neku oznaku, pogotovo kada su u pitanju fotorealistične slike.

5.1 Problem autorskih prava

Svi mi svakodnevno koristimo slike za razmenjivanje ideja i vizuelna reprezentacija sadržaja je sada više nego ikada dostupna svima nama. Pored lične upotrebe, slike su u savremenom kapitalističkom društvu postale sredstvo promovisanja usluga i proizvoda (naslovnice knjiga, bilbordi itd.). Generatori slika svima nama dozvoljavaju da kreiramo potpuno nove slike i da ih koristimo u komercijalne svrhe, ali onda se javlja pitanje ko polaže autorska prava nad tim generisanim slikama. Da li osoba koja je okucala tekst i pritisnula dugme za generisanje, da nije algoritam koji je izgenerisao sliku ili pak kompanija koja je kreirala algoritam?

Takođe, generatori slike imaju mogućnost da oponašaju umetnički stil mnogih poznatih umetnika. S jednim klikom korisnik može da izgeneriše hiljade slika u stilu Leonarda Da Vinčija. Da li je u tom slušaju korisnik plagirao životno delo slikara? Generator slika ne samo da uzima inspiraciju iz jedne slike umetnika, već kreira slike na osnovu njegovog celokupnog životnog dela.

Na prvu loptu ovo nam zvuči kao očigledan plagijat, ali moramo sagledati širu sliku. Kao što ljudskog biće dobija inspiraciju iz svega što ga okružuje, tako i veštačka inteligencija dobija inspiraciju iz kolekcija podataka koje analizira. U apstraktnom smislu, način na koji čovek i veštačka inteligencija kreiraju slike je dosta sličan. Oboje imaju neku početnu ideju koja proizašla iz informacija koje su dobili na njima prirodan način. Čovek ih dobija preko svojih čula, dok veštačka inteligencija ih, trenutno, dobija u binarnom zapisu na nekom medijumu za skladištenje podataka. Polazeći iz ideje, oboje biraju korake koji ih vode prema tome da završnim proizvod bude njihova lična kreacija. Jedina razlika između čoveka i veštačke inteligencije je činjenica da je veštačka inteligencija mnogo brža i efikasnija u tom procesu. Može se reći da je veštačka inteligencija umnogome sposobnija od čoveka.

Ukoliko bismo živeli u svetu gde je čovek podjednako inteligentan kao veštačka inteligencija ne bi nam bilo ni čudno ni pogrešno da čovek može da kreira s tolikom preciznošću i brzinom. To što je veštačka inteligencija trenutno samo jedno polje na veb stranici ne treba da nas zavara da će veštačka inteligencija jednog dana, vrlo verovatno, sama stvarati i evoluirati.

Generatori slika stvarno kreiraju problem celokupnoj kreativnoj industri-

ji, ali treba na veštačku inteligenciju posmatrati kao na još jednu alatku koju umetnik može da koristi. Možda upravo uz pomoć veštačke inteligencija umetnost može postati brža, dostupnija i otvorenija [17].

6 Zaključak

Dakle, znamo šta je veštačka inteligencija, kako se ona može upotrebiti (ali i zloupotребiti) za generisanje slika, kako radi, kao i par naših omiljenih primera. Naravno, i dalje postoje mnoga ograničenja, ali i mogućnosti za napredak u oblasti AI-a. Jedna od tih oblasti je generisanje videa iz unetog teksta na čemu radi poznata kompanija Meta⁹.

Oduprećemo se želji da vodimo debatu o tome kako će sve ovo uticati na kreativnu industriju. Međutim, istorija nam sugerise da pojava novih alata ima tendenciju da proširi definiciju umetnosti, odnosno *"umetnost nije mrtva, samo je mašinski generisana"* [18].

⁹Meta, nekada Facebook, jedna je od vodećih svetskih biznis magnata u oblasti razvoja veštačke inteligencije i tehnologije.

Literatura

- [1] Schank, R. C. (1987). What Is AI, Anyway?. AI Magazine, 8(4), 59.
<https://doi.org/10.1609/aimag.v8i4.623>
- [2] Dostupno na adresi

<https://www.cse.unsw.edu.au/~cs9417ml/RL1/introduction.html>
- [3] Dostupno na adresi

<https://www.youtube.com/watch?v=yTAMrHVG1ew&t=53s/>
- [4] Stanislav Frolov, Tobias Hinz, Federico Raue, Jörn Hees, Andreas Dengel, Adversarial text-to-image synthesis: A review. Dostupno na adresi
<https://www.sciencedirect.com/science/article/pii/S0893608021002823>
- [5] Scott Reed, Zeynep Akata, Xincheng Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee, Generative Adversarial Text to Image Synthesis. Dostupno na adresi <http://proceedings.mlr.press/v48/reed16.pdf/>
- [6] Computerphile, Generative Adversarial Networks (GANs) - Computerphile. Dostupno na adresi
<https://www.youtube.com/watch?v=Sw9r8CL98N0/>
- [7] Computephile, How AI Image Generators Work (Stable Diffusion / Dall-E) - Computerphile. Dostupno na adresi
<https://www.youtube.com/watch?v=1CIpzeNxIhU/>
- [8] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen, Hierarchical Text-Conditional Image Generation with CLIP Latents. Dostupno na adresi <https://cdn.openai.com/papers/dall-e-2.pdf/>
- [9] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, Mark Chen, GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. Dostupno na adresi <https://arxiv.org/abs/2112.10741/>
- [10] OpenAI, CLIP: Connecting Text and Images. Dostupno na adresi
<https://openai.com/blog/clip/>

- [11] AssemblyAI, How does DALL-E 2 actually work? Dostupno na adresi <https://www.youtube.com/watch?v=F1X4fHzF4mQ&t=359s/>
- [12] AssemblyAI, Diffusion models explained in 4-difficulty levels. Dostupno na adresi <https://www.youtube.com/watch?v=yTAMrHVG1ew&t=53s/>
- [13] Knjiga je besplatna i dostupna na adresi: <https://dallery.gallery/the-dalle-2-prompt-book/>
- [14] Rune Klingenberg Hansen, AI Image Generator: This Is Someone Thinking About Data Ethics. Dostupno na adresi <https://dataethics.eu/ai-image-generator-this-is-someone-thinking-about-data-ethics/>
- [15] DALL-E 2. Dodatne informacije dostupne na adresi <https://openai.com/dall-e-2/>
- [16] DALL-E 2 Content policy. Dostupno na adresi <https://labs.openai.com/policies/content-policy>
- [17] Paul Ford, Dear Artists: Do Not Fear AI Image Generators. Dostupno na adresi <https://www.wired.com/story/artists-do-not-fear-ai-image-generators/>
- [18] Guido Appenzeller, Matt Bornstein, Martin Casado, Yoko Li, *Art isn't dead, it's just machine-generated*