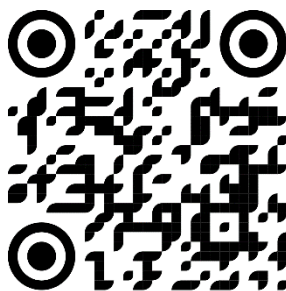


**PANEVROPSKI UNIVERZITET APEIRON, BANJA LUKA
FAKULTET INFORMACIONIH TEHNOLOGIJA**

Redovne studije
Smjer : “Inženjering Informacionih Tehnologija“

Predmet:
KOMPJUTERSKA GRAFIKA U DIGITALNOM DOBU

Ukrštanje viještačke inteligencije i kompjuterske grafike
(istorijski pregled i budući izgledi)
(Seminarski rad)



Predmetni nastavnik
Prof. dr Nedim Smailović

Student: Pavlović Ivan
Br. Indeksa : 92-20/RITP-S

Banja Luka, april 2024.

SADRŽAJ

UVOD	3
1. Viještačka inteligencija	4
1.1. Istorija viještačke inteligencije	4
1.2. Računarska nauka	6
1.3. Rođenje viještačke inteligencije (1941 – 1956).....	6
1.3.1. Turing-ov test.....	8
1.4. Savremena viještačka inteligencija (1993 -).....	8
1.4.1. Prekretnice u viještačkoj inteligenciji i Moore-ov zakon	8
2. Viještačka inteligencija u kompjuterskoj grafici	10
2.1. Evolucija viještačke inteligencije u generisanju video sadržaja	10
2.1.1. Porast AI-generisanih lažnih video zapisa	12
2.1.2. Neistraženo područje viještačke inteligencije u generisanju video sadržaja	13
2.2. Evolucija viještačke inteligencije u generisanju slika	14
2.2.1. Limitacije GAN tehnologije	16
2.2.2. DALL-E	17
2.2.3. CLIP	18
2.2.4. BigSleep.....	20
2.2.5. VQGAN-CLIP	21
2.2.6. Modeli difuzije.....	22
2.2.7. Prednosti i mane CLIP difuzije.....	24
2.2.8. Groznica generisanja slika pomoću viještačke inteligencije	25
ZAKLJUČAK	29
POPIS SLIKA	30
CITATNI IZVORI	31

UVOD

U današnjem digitalnom dobu, viještačka inteligencija je postala neizostavan dio naše svakodnevnice, transformirajući način na koji živimo, radimo i komuniciramo.

U ovom seminarkom radu prikazujem razvoj viještačke inteligencije, od njenih ranih koraka pa do savremenih dostignuća, sa posebnim fokusom na njenu primjenu u kompjuterskoj grafici. Proučavajući samu evoluciju viještačke inteligencije, njenih mogućnosti generisanja video sadržaja kao i u generisanju slikovnih sadržaja.

U prvom dijelu seminarskom rada, analiziramo istoriju viještačke inteligencije, tehnološke inovacije i prekretnice koje su oblikovale razvoj viještačke inteligencije. Unutar seminarskor rada, također se proučava Turingov test, Mooreov zakon i ostali faktori koji su doprinjeli stvaranju temelja savremene viještačke inteligencije.

Zatim, fokusiramo se na ulogu viještačke inteligencije u kompjuterskoj grafici, evoluciji viještačke inteligencije u generisanju video sadržaja, analiziranju porasta lažno generisanih video sadržaja pomoću viještačke inteligencije i neistražena područja ovog polja. Također, istražuje se napredak u generisanju slika, pregledavajući najnovije tehnologije poput GAN-a, DALL-E, CLIP, kao i njihove prednosti i mane.

1. Vještačka inteligencija

1.1. Istorija vještačke inteligencije

„Istorija vještačke inteligencije (engl. Artificial Intelligence) započela je u antičkom dobu, sa mitovima, pričama i glasinama o vještačkim bićima koja su obdarena inteligencijom ili svijesti od strane majstorskih zanatlija. Sami početci modernog Ai-a započinjane su od strane filozofa koji su pokušali da opišu proces ljudskog razmišljanja kao mehaničku manipulaciju simbola. Ovaj rad je kulminirao nastankom prvog digitalnog računara u 1940-im godinama, mašine zasnovane na apstraktnoj suštini matematičkog razmišljanja. Prvi računari 1940-ih godina i kasnije su inspirisali nekolino naučnika da ozbiljno razmišljaju o mogućnosti izgradnje elektronskog mozga“ [1].



Slika 1 - Vještački mozak

„Polje istraživanja vještačke inteligencije je osnovano na radionici održanoj na kampusu Dartmouth koledžu u SAD-u tokom ljeta 1956. godine. Oni koji su prisustvovali istraživanju su postali lideri u istraživanju vještačke inteligencije za naredne decenije. Mnogi od njih su predviđali da će mašina inteligentna kao čovjek postojati u roku od jedne generacije, i dodijeljeno im je na milione američkih dolara da svoje vizije pretvore u stvarnost“ [1].

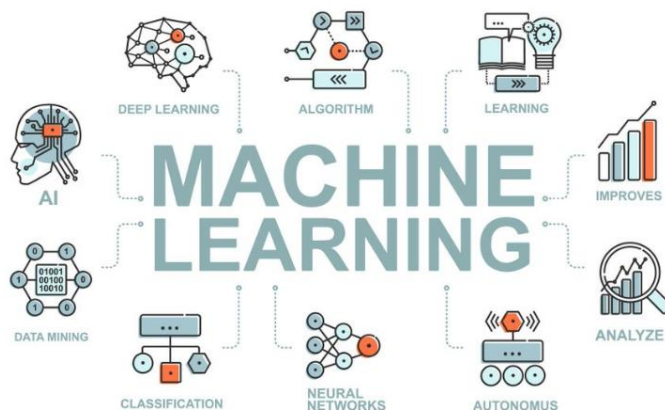


Slika 2 - Logo "Dartmouth" koledža

„Na kraju postalo je očigledno da su istraživači drastično podcijenili težinu projekta. Godine 1974, kao odgovor na kritike od strane James Lighthill-a i stalni pritisak kongresa, vlade SAD-a i Velike Britanije su prestale finansiranje istraživanja viještačke inteligencije, a teške godine koje su uslijedile nakon toga su također poznate kao zima viještačke inteligencije (engl. Ai Winter).

Sedam godina kasnije, vizionarska inicijativa Japanske vlade inspirisala je vlade i industriju da obježbede resurse za daljnje istraživanje viještačke inteligencije, ali do kraja 1980-tih investitori su postali razočarani i ponovo povukli finansiranje.

Investiranje i interes za viještačku inteligenciju su ponovo pokrenute 2020-ih kada je mašinsko učenje uspješno primjenjeno na mnoge probleme u akademskoj sferi i industriji zahvaljujući novim metodama, primjeni moćne računarske opreme i prikupljanju ogromnih skupova podataka“ [1].

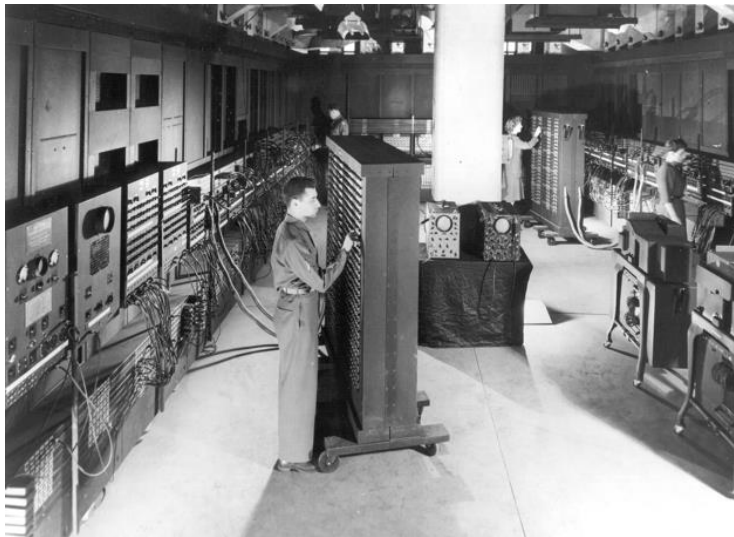


Slika 3 - Mašinsko učenje i njeni pod-tipovi

1.2. Računarska nauka

„Računarske mašine su bile dizajnirane ili pravljene u antičko doba i tokom istorije od strane mnogo ljudi, uključujući Gottfried Leibniz, Joseph Marie Jacquard, Charles Babbage, Percy Ludgate, Leonardo Torres Quevedo, Vannevar Bush, i ostali. Ada Lovelace je spekulisala da je Babbageova mašina „misleća ili ... zaključujuća mašina“, ali je upozorila da je poželjno čuvati se od mogućnosti preuveličavanja ideja koje se javljaju u vezi sa moćima.

Prvi moderni računari bili su masivne mašine Drugog svjetskog rata kao što su Konrad Zuse Z3, Colossus, ENIAC i mnogi drugi. ENIAC je bio zasnovan na teorijskom temelju postavljenom od strane Alan Turing-a i razvijen od strane John Von Neumann-a i kasnije se pokazao najuticajnijim“ [1].



Slika 4 - Fotografija Američke vojske u korištenju ENIAC-a

1.3. Rođenje viještačke inteligencije (1941 – 1956)

„Najranije istraživanja o mašinama koje razmišljaju bila su inspirisana spojem ideja koje su postale uobičajene krajem 1930-tih, 1940-ih i početkom 1950-tih godina. Nedavna istraživanja u neurologiji pokazala su da je mozak električna mreža neurona koji se aktiviraju u impulsima sve ili ništa“ [1].

„Norbert Winer-ova kibernetika opisala je kontrolu i stabilnost u električnim mrežama. Claude Shannon-ova teorija informacija opisala je digitalne signal. Alan Turing-ova teorija računanja pokazala je da se bilo koji oblik računanja može opisati digitalno. Uska veza između ovih pojmova, sugerisala je da bi bilo moguće konstruisati „elektronski mozak““ [1].



Slika 5 - IBM 702 - Prvi računar korišten od strane prvih istraživača vještačke inteligencije

„U 1940-im i 50-im godinama, nekolicina naučnika iz različitih oblasti (matematike, psihologije, ekonomije i političkih nauka) istraživalo je nekoliko pravaca istraživanja koji će biti ključni za kasnija istraživanja vještačke inteligencije.

Alan Turing, koji je razvio teoriju računanja, bio je među prvima koji su ozbiljno istraživali teorijsku mogućnost „mašinske inteligencije“. Polje istraživanja „vještačke inteligencije“ osnovano je kao akademska disciplina 1956. godine“ [1].



Slika 6 - Portret Alan Turing-a

1.3.1. Turing-ov test

„Alan Turing je razmišljao o mašinskoj inteligenciji najkasnije od 1941. godine, kada je kružio rad o mašinskoj inteligenciji koji bi mogao biti najraniji rad u oblasti viještačke inteligencije – iako je sada izgubljen. 1950. godine Turing je objavio prekretnički rad „Computing Machinery and Intelligence“ u kojem je spekulisao o mogućnosti stvaranja mašina koje razmišljaju, a rad je predstavljao njegov koncept koji je danas poznat kao Turingov test.

Alan je uočio da je „razmišljanje“ teško definisati i osmislio je svoj čuveni Turingov test. Ako bi mašina mogla voditi razgovor koji je bio neodvojiv od razgovora s ljudskim bićem, tada je bilo razumno reći da mašina „razmišlja“.

Ova pojednostavljena verzija problema omogućila je Turing-u da argumentuje da je „razmišljajuća mašina“ barem moguća. Turingov test je bio prvi ozbiljan prijedlog u filozofiji viještačke inteligencije“ [1].

1.4. Savremena viještačka inteligencija (1993 -)

„Oblast viještačke inteligencije, sada starije od pola vijeka, konačno je postiglo neke od svojih najstarijih ciljeva. Počelo je uspješno da se koristi u cijeloj tehnološkoj industriji, iako nešto manje vidljivo. Dio uspjeha je rezultat povećane računarske snage, a dio je postignut fokusiranjem na specifične izolovane probleme i njihovih rješavanjem uz najviše standarde naučne odgovornosti.

Unutar ove oblasti nije postignuta saglasnost o razlozima zbog kojih viještačka inteligencija nije ispunila san o inteligenciji na nivou ljudskog bića koji je zauzeo maštu svijeta 1960-tih godina“ [1].

1.4.1. Prekretnice u viještačkoj inteligenciji i Moore-ov zakon

„Na 11. maj 1997. godine, Deep Blue postao je prvi računarski sistem za igranje šaha koji je pobijedio tadašnjeg svjetskog šampiona u šahu, Garry Kasparov-a. Super-kompjuter je bio

specijalizovana verzija okvira proizvedenog od strane IBM-a i bio je sposoban da obradi dva puta više poteza u sekundi nego tokom prvog meča (koji je Deep Blue izgubio), navodno 200.000.000 poteza u sekundi“ [1].



Slika 7 - "Deep Blue" super-komputer

„U 2005. godini, robot sa Univerziteta Stanford je pobijedio na DARPA Grand Challenge takmičenju vozeći autonomno 131 kilometara duž staze u pustinji koja nije bila unaprijed pripremana. Dvije godine kasnije, tim sa Univerziteta Carnegie Mellon je također pobijedio na DARPA Urban Challenge takmičenju autonomno prolazeći 55 kilometara kroz urbano okruženje uz poštovanje saobraćajnih opasnosti i propisa.

Ovi uspjesi nisu rezultat revolucionarne nove paradigme, već su uglavnom posljedica temeljne primjene inženjerskih vještina i ogromnog povećanja brzine i kapaciteta računara do devedesetih godina. Zapravo, računar Deep Blue bio je deset miliona puta brži od Ferranti Mark 1 računara koji je osmislio Christopher Strachey da igra šah.

Ovo dramatično povećanje se mjere Moore-ovim zakonom, koji predviđa da će se brzina i kapacitet memorije računara udvostručiti svake dve godine, kao rezultat udvostručavanja broja metal-oksid-poluprovodničkih tranzistora svake dve godine“ [1].

2. Viještačka inteligencija u kompjuterskoj grafici

2.1. Evolucija viještačke inteligencije u generisanju video sadržaja

„Brzi napredak viještačke inteligencije, revolucionisao je različite industrije, a jedna od najzapaženijih transformacija desila se u oblasti generisanja video sadržaja. Kroz godine, viještačka inteligencija je progresivno evoluirala, oblikujući način na koji se kreiraju video materijali, unapređujući kreativni proces i pomjerajući granice vizuelne priče“ [2].



Slika 8 - Primjer stvarnog videa pretvoren u animaciju pomoću AI-a

„Korijeni viještačke inteligencije u generisanju video sadržaja mogu se pratiti sve do samog početka računarske grafike i algoritama. Tokom 1960-ih i 1970-ih godina, računarski generisana grafika (CGI) počela je da se pojavljuje, omogućavajući kreiranje jednostavnih vizuelnih efekata i grafika. Međutim, ovi rani naponi bili su ograničeni računarskom snagom i nedostajala im je složenost koju će kasnije viještačka inteligencija unijeti u ovu oblast.

Prekretnica nastaje sa pojavom neuronskih mreža tokom 1980-ih godina. Neuronske mreže, inspirisane strukturom ljudskog mozga, omogućile su računarima da procesuiraju i prepoznaju obrasce u podacima. Ovo je označilo početak nove ere u generisanju video sadržaja uz pomoć viještačke inteligencije, omogućavajući kreiranje složenijeg i nijansiranjem vizuelnog sadržaja.

Prava transformacija viještačke inteligencije u generisanju video sadržaja nastupila je sa usponom dubokog učenja (engl. deep learning) i neuronskim mrežama tokom 2010-ih godina. Algoritmi dubokog učenja pokazivali su izuzetnu sposobnost učenja iz velikih skupova podataka i generisanja složenih vizuala. Ovaj proboj je otvorio put ka značajnim napretcima u obradi video materijala, uključujući detekciju objekata, segmentaciju slika i prepoznavanje dešavanja.

Model „Text-to-Video“ je model mašinskog učenja koji kao ulaz uzima opis prirodnog jezika i proizvodi video koji odgovara tom opisu. Ova tehnologija se ostvaruje pomoću neuronskih mreža za model sekvence na sekvencu sa konektorom konvolutivnih neuronskih mreža kodiranja i dekodiranja svakog piksela, svakog frejma, kreirajući video korištenjem dubokog učenja“ [2].

U ranim iteracijama „Text-to-Video“ modela, korisnik je unio opisnu informaciju „Will Smith eating spaghetti“, što je rezultiralo video sadržajem koji je nekima čak izgledao uznemiravajuće ili smiješno. Video demonstraciju možemo pogledati na slijedećem linku:

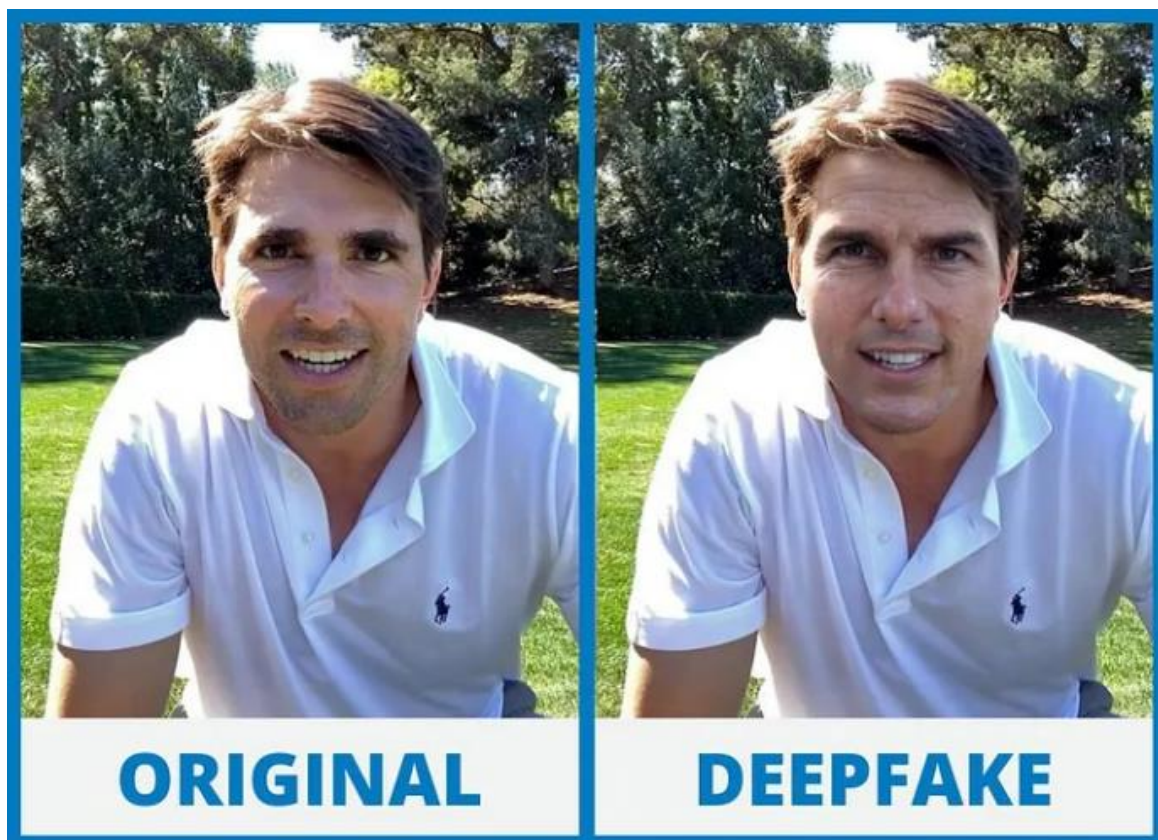
https://www.youtube.com/watch?v=XQr4Xklqzw8&ab_channel=RobotNamedRoy



Slika 9 - Prvi primjeri korištenja "Text-to-Video" tehnologije koja je malo uznemiravajuća

2.1.1. Porast AI-generisanih lažnih video zapisa

„Iako je viještačka inteligencija dizajnirana da otključa ogroman potencijal u kreiranju video sadržaja, istovremeno je izazvala etičke zabrinutosti, posebno u oblasti „Deep Fakes“ tj. lažnih video zapisa. Lažni video sadržaji su digitalno manipulirani video zapisi koji prikazuju pojedince kako govore ili rade stvari koje zapravo nisu radili. Dok algoritmi viještačke inteligencije postaju sve sofisticiraniji, raste prijetnja dezinformacija i manipulaciji“ [2].



Slika 10 - Primjer "Deep Fake" sadržaja

„Studij koji je sprovedla kompanija za prijetnje viještačkom inteligencijom, Sensity, otkrila je da se broj „Deep Fake“ video zapisa na internetu udvostručio za samo devet mjeseci, što naglašava brzo širenje ove tehnologije. To ističe hitnu potrebu za pouzdanim mehanizmima detekcije i etičkim smjernicama kako bi se umanjili potencijalni negativni uticaji sadržaja generisanog viještačkom inteligencijom“ [2].

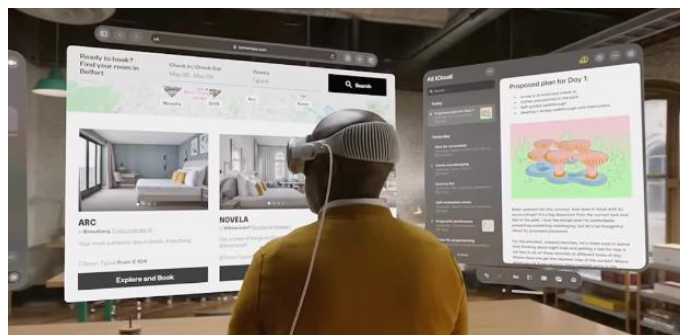
2.1.2. Neistraženo područje viještačke inteligencije u generisanju video sadržaja

„Gledajući unaprijed, evolucija viještačke inteligencije u generisanju video sadržaja ne pokazuje znakove usporavanja. Dok istraživači nastavljaju da guraju granice AI algoritama, možemo očekivati još revolucionarnih razvoja. Transfer stila u realnom vremenu, generisanje personalizovanog sadržaja i unapređeni alati za editovanje videa samo su neka od područja koja nose ogroman potencijal za rast.



Slika 11 - AppleVision Pro

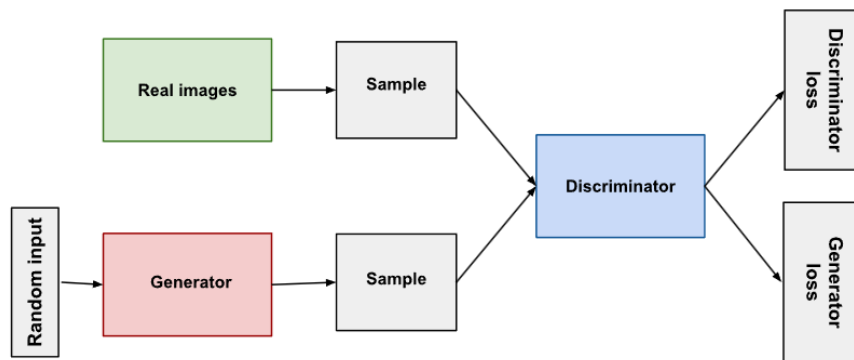
Evolucija AI-a u generisanju video sadržaja je dokaz ljudske inventivnosti i tehnološkog napretka. Od skromnih početaka u osnovnim algoritmima do pojave neuronskih mreža, dubokog učenja, GAN-ova i dalje, AI je preoblikovao način na koji se stvaraju i konzumiraju video sadržaji. Iako izazovi poput deepfake-ova postavljaju etničke dileme, također služe kao podsjetnik o važnosti odgovornog razvoja i korištenja AI tehnologije. Dok prolazimo kroz ovaj dinamični pejzaž, jasno je da je istorijski uticaj AI-a na generisanje video sadržaja samo početak uzbudljivog putovanja u budućnost“ [2].



Slika 12 -- Tačka gledišta AppleVision Pro korisnika

2.2. Evolucija viještačke inteligencije u generisanju slika

„Najraniji pokušaji u generisanju slika uz pomoć viještačke inteligencije su se desili još 1970-tih godina, decenijama nije bilo mnogo napretka u toj oblasti. Raspoloživa računarska snaga i količina podataka bili su ograničeni, a algoritmi su bili previše jednostavni i kruti da bi se nosili sa složenijim i realističnijim slikama. Međutim, ova situacija je počela da se mijenja sa pojavom dubokog učenja i konvolutivnih neuronskih mreža, koje su čak poslužile kao osnova za **Generativne Suparničke Mreže (GAN)**.



Slika 13 - Struktura GAN mreže

GAN-ovi su označili značajan napredak u oblasti slika generisanih viještačkom inteligencijom. Ova neuronska mreža je bila razvijena u 2014-oj godini od strane Ian Goodfellow-a i njegovih kolega sa Univerziteta u Montrealu. Sastoji se od dve neuronske mreže: generatora i diskriminatora, koje se obučavaju u smijenjujućim periodima. Generator mreža uči da generiše sintetičke podatke koji imitiraju distribuciju stvarnih podataka, dok diskriminator mreža uči da razlikuje između sintetičkih podataka generisanih od strane generatora i stvarnih podataka.



Slika 14 - Generator i diskriminator u praksi

Proces obuke je iterativan, pri čemu generator pokušava da proizvede sintetičke podatke koji mogu pretvarati diskriminatora, dok se diskriminator trudi da poboljša svoju sposobnost razlikovanja između stvarnih i lažnih podataka.

Krajnji rezultat je generator koji može proizvesti sintetičke podatke koji su vizualno ili auditivno neodvojivi od stvarnih. Ovaj opšti pristup može se koristiti ne samo za sintezu slika već i za druge zadatke, kao što su transfer stila ili augmentacija podataka, i može se primijeniti i izvan vizuelnog domena, na primjer, u generisanju zvuka, poput sintetizovanja muzike.

Ovaj pristup se pokazao veoma uspješnim u generisanju visoko-kvalitetnih slika i koristi se u različitim aplikacijama, kao što su za generisanje slika lica, prejaža, objekata itd. Osim toga, GAN-ovi su prilagođeni i za druge zadatke, uključujući generisanje detaljnih slika iz grubih skica, transfer stila sa jedne slike na drugu ili zamjenjivanjem dijelova slike željenim objektom“ [3].

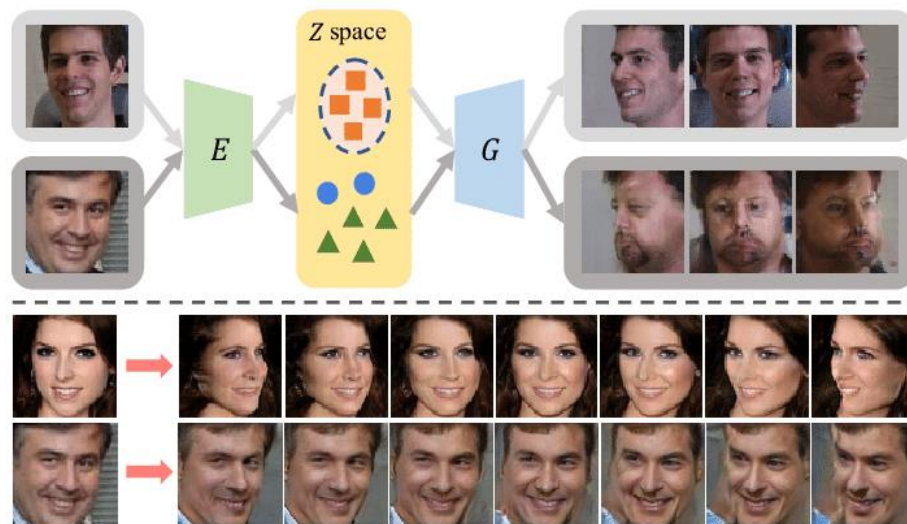


Slika 15 - Automobil koji ne postoji, kreiran pomoću GAN tehnologije

2.2.1. Limitacije GAN tehnologije

„Iako, GAN-ovi su pokazali impresivne rezultate u generisanju realističnih slika, postoji nekoliko ograničenja ove tehnologije. Jedno od tih ograničenja je nestabilna priroda procesa obuke. Protivnička struktura mreže može dovesti do kolapsa, gdje generator mreža proizvodi samo ograničen skup slika koji ne pokriva cijeli raspon mogućih rezultata, što rezultira nedostatkom raznolikosti u izlaznim slikama.

Preprilagođavanje je još jedan problem kojem GAN-ovi mogu biti posebno podložni, gdje generator mreža pamti skup za obuku umjesto da ga generalizuje kako bi generisala nove slike. Dodatno, GAN-ovi mogu imati poteškoća u generisanju visoko-kvalitetnih slika visoke rezolucije zbog složenosti računarskih operacija koje su uključene, često zahtijevajući značajnu količinu vremena i računarske snage za obuku.



Slika 16 - Limitacije GAN tehnologije

Osim toga, iako postoje primjeri GAN-ova koji prevode tekst u sliku ili modele koji vrše translaciju slike u sliku, većina implementacija pruža ograničenu kontrolu nad generisanim izlazom, što otežava proizvodnju specifičnih objekata ili promjenu stila generisanih slika bez značajne dodatne obuke ili manipulacije“ [3].

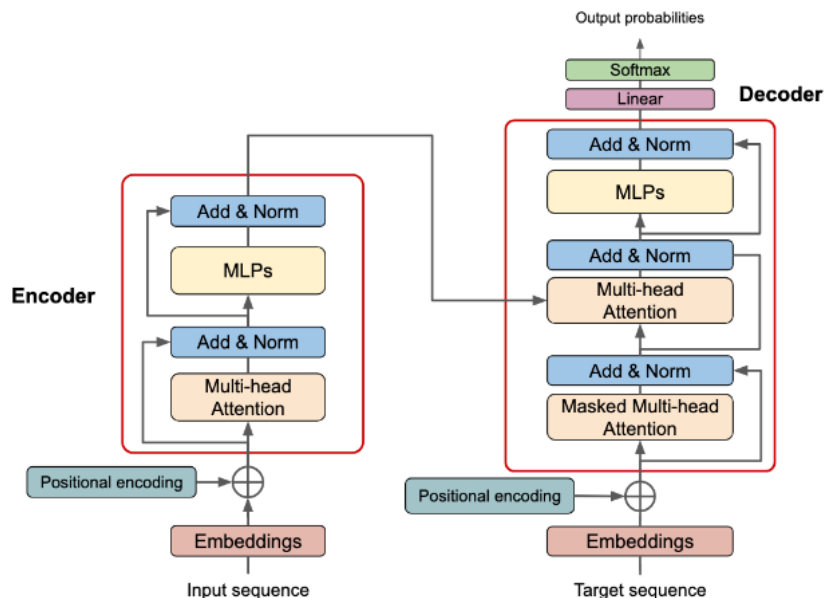
2.2.2. DALL-E

„Da bi prevazišli ova ograničenja, istraživači su nastavili da istražuju nove tehnike i arhitekture za generisanje slika. Jedan takav primjer je DALL-E, razvijen od strane OpenAI-a i objavljen 5. januara 2021. godine. Model koristi generativni pre-trenirani transformer, koji se dalje bazira na ranijem modelu transformatora. Oba su originalno razvijena za upotrebu u obradi prirodnog jezika.



Slika 17 - Logo OpenAI-a i logotip DALL-E softvera

Model transformera je arhitektura neuronske mreže zasnovana na mehanizmu samopažnje. U tradicionalnim neuronskim mrežama, svaki ulazni element se obrađuje nezavisno, što može dovesti do poteškoća u modeliranju zavisnosti na dugim udaljenostima. Mehanizam pažnje omogućava modelu da selektivno fokusira različite dijelove ulaznog niza, omogućavajući mu da uhvati složene odnose između riječi.



Slika 18 - Arhitektura Transformera u DALL-E

Model transformera sastoji se od enkodera i dekodera, koji oba sadrže više slojeva samopažnje i neuronskih mreža unaprijeđenog propusnog opsega. Visoko je paralelizovan i postigao je rezultate koji su vodeći u mnogim zadacima obrade prirodnog jezika.

Na osnovu te arhitekture, generativni pre-trenirani transformeri su originalno razvijeni 2018-te godine od strane OpenAI-a. GPT funkcioniše tako što trenira veliku neuronsku mrežu na ogromnim količinama tekstualnih podataka, poput knjiga, članaka i veb stranica. Model koristi proces nazvan nenadgledano učenje kako bi identifikovao obrasce i odnose u podacima, ne bivajući eksplicitno upućen.



Slika 19 - Slika generisana od strane DALL-E koristeći input "fotelja u obliku avokada"

Model DALL-E je pokazao izvanredne rezultate u više zadataka generisanja slika, baziranih na tekstualnim ulazima. Osim što jednostavno generiše uzorke slika različitih objekata koje je vidio tokom treninga, model posjeduje sposobnost integrisanja različitih ideja i spajanja nespojenih koncepata na ubjedljiv način, čak i generišući objekte koji su vjerovatnoćom neobični za postojanje u fizičkom svijetu“ [3].

2.2.3. CLIP

„Još jedan važan korak u sitezi slika pomoću viještačke inteligencije, objavljen zajedno sa DALL-E modelom, bila je tehnika nazvana Kontrasno pre-treniranje jezika-slike (CLIP). CLIP je model obučen na 400 miliona parova tekstualnih opisa i slika, pri čemu su opisi preuzeti sa interneta. Uključuje treniranje neuronske mreže na podacima kako slika, tako i teksta sa ciljem omogućavanja modelu da razumije odnos između ova dva modaliteta – kako se tekstualni opisi odnose na vizuelni sadržaj slika.

Tokom treninga, model uči da enkoduje ulazne slike i tekstove u zajednički vektorski prostor, gdje se sličnost između slike i njene odgovarajuće tekstualne deskripcije maksimizuje, dok se sličnost između slika i drugih irelevantnih tekstova minimizuje.

U DALL-E, CLIP je korišten za rangiranje generisanih slika po tačnosti opisa, kako bi se filtrirao početni skup slika i odabrao najprikladniji izlaz. Od tada, CLIP je postao popularan element koji se koristi u različitim rješenjima za sintezu slika, služeći kao filter i vodič u procesu generisanja slika.

Prvo open-source rješenje koje je koristilo ovaj model bio je „DeepDaze“. Razvijen u Januaru 2021. godine od strane Phil Wang-a, kombinovao je CLIP sa neuronskom mrežom implicitne reprezentacije nazvanom Siren. DeepDaze je postao popularan zbog svoje sposobnosti da proizvodi vizuelno impresivne i nadrealne slike koje često podsjećaju na sanjive pejzaže ili apstraktnu umjetnost“ [3].

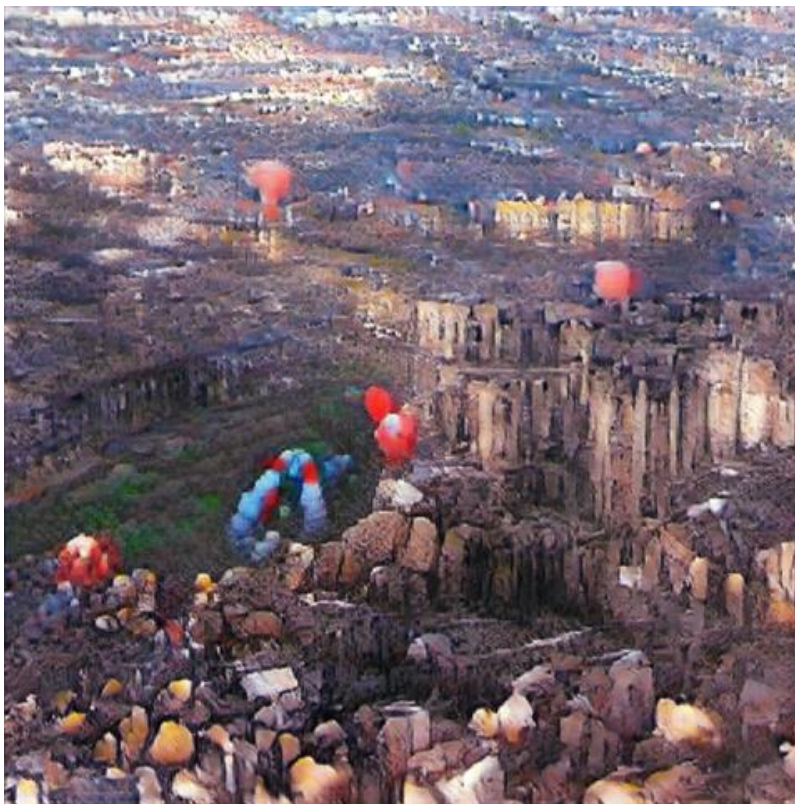


Slika 20 - "Putnik kroz vrijeme u gužvi" generisan od strane DeepDaze

2.2.4. BigSleep

„Nekoliko dana kasnije, isti programer, uz pomoć modela koje je objavio istraživač Ryan Murdock, razvio je još jedan generativni model dubokog učenja nazvan „BigSleep“. Model funkcionise tako što kombinuje CLIP sa BigGAN-om, sistemom koji su razvili istraživači u kompaniji Google, a koji koristi varijantu GAN arhitekture kako bi generisao slike visoke rezolucije iz vektora slučajnog šuma. BigSleep koristi izlaze iz BigGAN-a kako bi pronašao slike koje postižu visok rezultat na CLIP-u. Zatim model postepeno prilagođava ulaz šuma u generatoru BigGAN-a dok proizvede slike odgovaraju datom uputstvu.

Prema Ryan Murdock-u, BigSleep je prvi model koji je mogao generisati raznovrsni spektar koncepata i objekata visokog kvaliteta u rezoluciji 512x512 piksela. Dok su prethodni radovi, iako su proizvodili impresivne rezultate, često bili ograničeni na slike niže rezolucije i uobičajene objekte“ [3].



Slika 21 - "Baloni iznad ruševina grada" generisano pomoću BigSleep-a

2.2.5. VQGAN-CLIP

„Model BigSleep je inspirisao još jednu kombinaciju CLIP-GAN. Samo tri mjeseca kasnije, u aprilu 2021. godine, istraživači Katherine Crowson, razvila je model VQGAN-CLIP. VQGAN (engl. Vector Quantized Generative Adversarial Network) je varijacija arhitekture generativno-suprotstavljenih mreža. VQGAN je sposoban da generiše visoko kvalitetne slike kodiranjem slika kao diskretnih ulaza u kodnu knjigu, što omogućava efikasnije obučavanje i bolji kvalitet slika u poređenju sa tradicionalnim GAN modelima.



Slika 22 - Sadržaj generisan VQGAN-CLIP metodom

VQGAN-CLIP koristi multimodularni enkoder CLIP-a kako bi procijenio sličnost para (tekst, slika) i vraća ga kroz mrežu u latentni prostor (tj. prostor apstraktnih reprezentacija slika) generatora slika VQGAN-a. Ponavljajući ovaj proces iterativno, generisani kandidat za sliku se prilagođava dok postaje sličniji ciljnom tekstu.

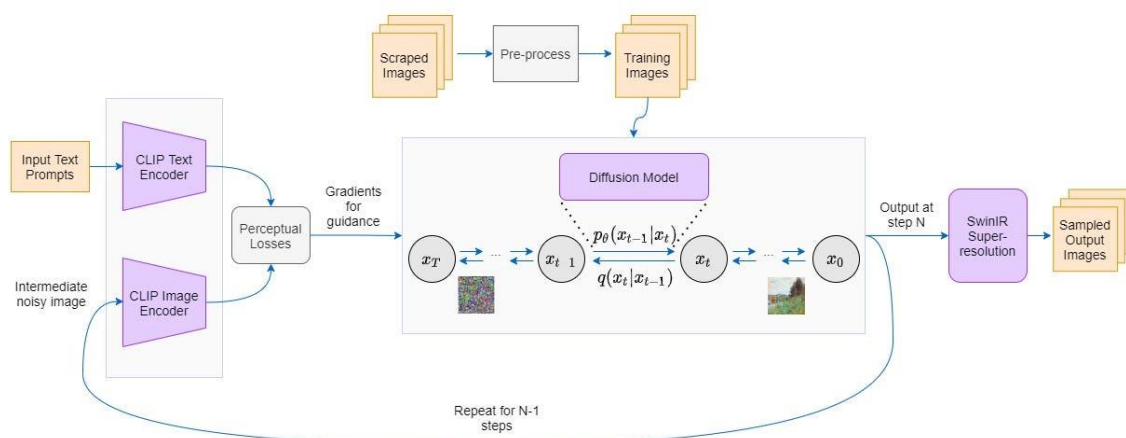
Model je omogućio ne samo generisanje slika, već i sposobnost manipulacije već postojećih slika, pružajući tekstualni opis promjene koja treba da se izvrši. Ova arhitektura predstavlja važan razvoj u modelima teksta-u-sliku, pružajući generisanje i manipulaciju visoko kvalitetnim slikama, semantičku tačnost između teksta, slike i efikasnosti, čak i pri generisanju neobičnog sadržaja“ [3].

2.2.6. Modeli difuzije

„Linija istraživanja koje se zasniva na povezivanju CLIP-a sa drugim arhitekturama nastavljena je. U junu 2021. godine, autor iza VQGAN-CLIP modela objavio je još jedan rad, kombinovanjem Contrastive Language-Image Pre-training model sa difuzionim algoritmom kako bi stvorio CLUP Guided Diffusion.

Difuzionni algoritmi su familija metoda za generisanje slika, zasnovanih na simulaciji ponašanja čestica koje difuziraju kroz sredstvo. Podaci o slici mogu se pretvoriti u uniformnu distribuciju dodavanjem slučajnog šuma. Difuzionni algoritmi postepeno razgrađuje strukturu slike dodavanjem šuma sve dok ne ostane samo slučajni šum.

U mašinskom učenju, cilj je naučiti proces reverzne difuzije – pronaći parametre za proces koji iterativno transformiše slučajni šum nazad u koherentnu sliku. Taj proces rekonstrukcije je uslovljen nekom vrstom signala vođenja“ [3].



Slika 23 - Arhitektura CLIP difuzije

U osnovi, algoritam radi na slijedeći način:

- Algoritam inicijalizuje podatke počevši sa signalom šuma koji se slučajno generiše – obično to je Gausova distribucija sa nultom sredinom vrijednošću i jediničnom varijansom
- Signal šuma se koristi kao početno stanje difuzionog procesa i čestice koje predstavljaju položaj piksela slike se uzrokuju iz te početne distribucije
- Simulira se difuzija čestica, modelirana kao niz vremenskih koraka, pri svakom od kojih se položaji čestica ažuriraju
- Proces je uslovljen vođenim signalom – ili slikom ili tekstualnim uputstvom
- Signal se koristi kako bi usmjerio difuziju ka generisanju slike koja odgovara željenim karakteristikama
- Koraci difuzije i uslovljavanje se ponavljaju iterativno za fiksni broj koraka
- Kako proces napreduje, čestice postaju koncentrisane u regionima koji odgovaraju vođenom signalu
- Kada se difuzicioni proces završi, konačni položaji čestica se transformišu u sliku koristeći dekodner neuronske mreže
- Opcionalno, mogu se preduzeti dodatni post-processing koraci, poput korekcije boje ili samog poboljšanja kvaliteta slike

„U CLIP Guided Diffusion-u, autor je koristio Guided Diffusion implementaciju difuzionog algoritma koji su napravili istraživači iz OpenAi-a, vođenu signalom koji pruža CLIP. U ovoj arhitekturi, CLIP model se koristi kako bi pronašao skup slika koje odgovaraju tekstualnom uputstvu i koristi ih kao vođenje za Guided Diffusion model. Korisnik također može pružiti dodatna ograničenja o željenoj slici, poput rezolucije slike ili paleti boja.

Zatim, Guided Diffusion model generiše niz vektora slučajnog šuma koji se postepeno usavršavaju kako bi generisali sliku koja odgovara vođenju koje je pružio CLIP. Model iterativno ažurira vektore slučajnog šuma uslovljavajući ih prethodnim vektorom šuma, vođenjem koje je pružio CLIP i slučajnim difuzionim šumom. Dobiveni vektori šuma zatim se transformišu u sliku koristeći dekodner neuronske mreže“ [3].

2.2.7. Prednosti i mane CLIP difuzije

„CLIP Guided Diffusion je pokazao sposobnost da generiše slike visokog kvaliteta sa realističnim teksturama i finim detaljima koji su teško postižni drugim generativnim modelima. Još važnije, on je također visoko kontrolabilan, što omogućava korisnicima da generišu slike koje se podudaraju sa specifičnim stilovima ili karakteristikama navedenim u tekstualnom uputstvu. Model se može koristiti za širok spektar zadataka generisanja slika, uključujući uređivanje slika, sinteza slika i manipulaciju slikama“ [3].



Slika 24 - Slika generisana pomoću CLIP Guided Diffusion-a na upit "Žena stoji u parku"

„Ovo otkriće je izazvalo eksploziju interesovanja za generisanje slika, šireći znanje o tehnologiji i na još veću publiku. Model Crowson-a uskoro je pratila Disco Diffusion – popularna sveska na Google kolaboraciji koja je koristila CLIP Guided Diffusion kako bi pružila veliki broj prilagodljivih opcija. Ovaj alat brzo je postao popularan među umjetnicima i entuzijastima za viještačku inteligenciju koji su sada imali još lakši i praktičniji pristup tehnologiji koja je ranije bila dostupna samo istraživačima u velikim tehnološkim kompanijama.

Jedna od mana difuzionog modela je ta što zahtijeva mnogo računarskih resursa za obuku i pokretanje, ali slijedeći razvoji bi trebali promijeniti tu činjenicu“ [3].

2.2.8. Groznica generisanja slika pomoću vještačke inteligencije

„Modeli vještačke inteligencije za generisanje slika pomoću teksta su u samo nekoliko godina mnogo napredovali, ali do 2022. godine su ostali prilično ograničeni. Mnogi raniji modeli su zahtijevali veliku računarsku snagu za rad, posebno tokom faze obuke. To je značilo da su često obučavani na manjim skupovima podataka, ostajući tako uglavnom zanimljivi samo istraživačima. Neki zanimljivi tehnološki alati bili su dostupni putem Google Notepad bilježnice, što je omogućavalo jednostavno izvršavanje koda na hardveru dostupno putem usluge u oblaku (engl. cloud). Međutim, ovaj metod je idalje zahtijevao određenu tehničku kompetentnost, što je ograničavalo popularnost osnovne tehnologije.

To je počelo da se mijenja 2022. godine, kada su počele da se pojavljuju nove aplikacije koje pružaju praktična učenja za modele zasnovane uglavnom na nekoj implementaciji difuzionog modela.

U martu 2022. godine, nezavisni istraživački laboratorijum MidJourney lansirao je otvorenu beta verziju svoje aplikacije, pružajući pristup visoko-kvalitetnom, umjetnički stilizovanom generisanju slika puta četa dostupne aplikacije za komunikaciju, zvanu Discord. Iako usluga zahtijeva plaćanje za kontinuiranu uslugu, idalje omogućava novim korisnicima generisanje ograničenog broja slika besplatno“ [3].

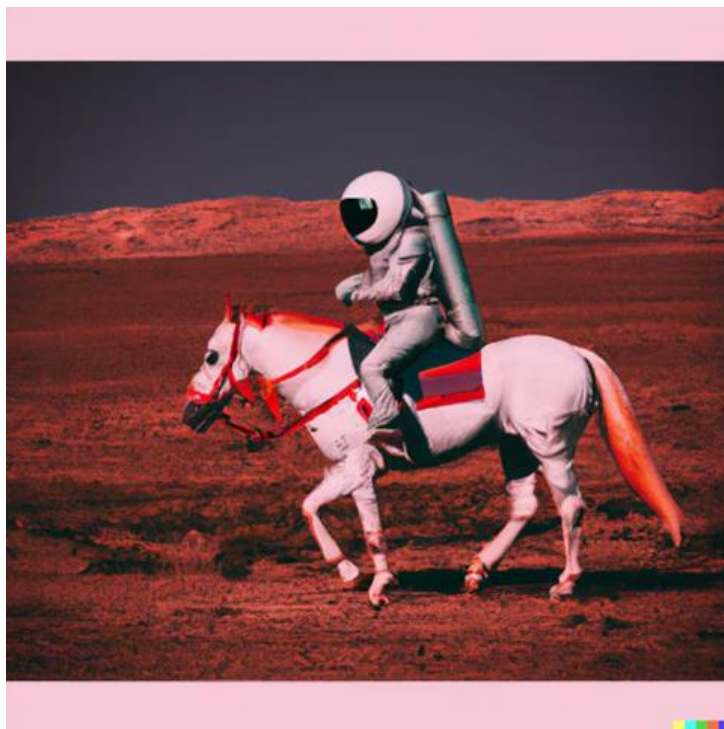


Slika 25 - "Mehanička ptica" kreirana četvrtom verzijom MidJourney softvera

„Uskoro nakon toga, u aprilu, još jedna velika web usluga za sintezu slika je lansirana. Prethodno dostupan samo ograničenoj publici, DALL-E 2 kompanije OpenAI je otvoren za javnost. Kao i MidJourney, zahtijeva kupovinu kredita za trošenje na generisanje slika ali nudi mali broj neakumulativnih kredita svakog mjeseca besplatno.

DALL-E 2 je kreiran obukom neuronske mreže na parovima slika i tekstualnih opisa. On može da razumije objekte kombinovanjem tekstualnih opisa sa slikama i također razumije odnose između objekata. Kao i mnogi drugi modeli najnovije generacije, koristi difuzicioni algoritam u procesu sinteze slika.

DALL-E 2 koristi CLIP kako bi naučio veze između tekstualne semantike opisa slike i njihovih vizuelnih reprezentacija, enkodirajući sliku u prostor prezentacije. Zatim se koristi modifikovana verzija jednog od prethodnih modela kompanije OpenAI, nazvada GLIDE, da inverzno procesuje enkodiranje slike kako bi dekodirao ugnježdavanje slike stvorene od CLIP-a. Ovaj metoda pomaže u kreiranju varijanti datog ulaza koje, iako su različite od originalne slike ili tekstualnog unosa, mogu idalje proizvesti rezultate koji obuhvataju karakteristične osobine pruženog unosa“ [3].

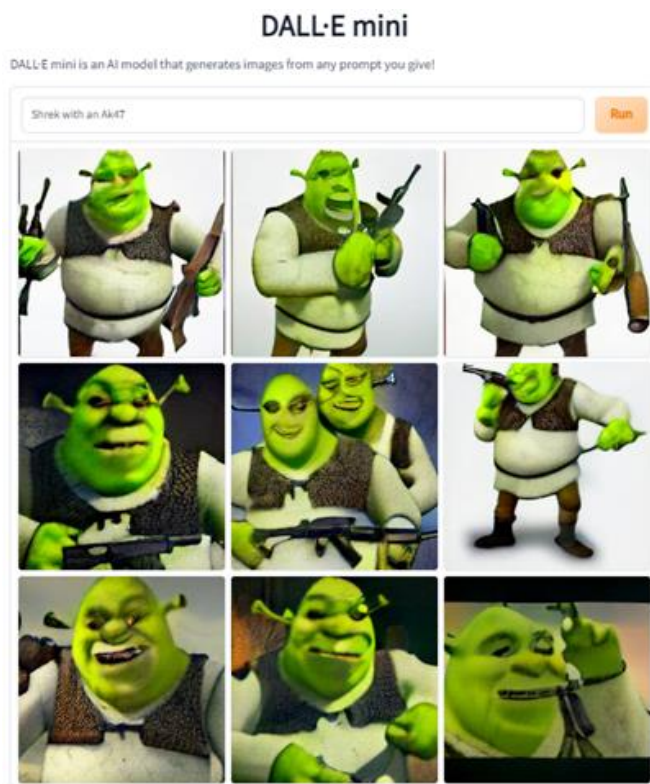


Slika 26 - "Astronaut koji jaše konja u foto-realističnom stilu" generisano pomoću DALL-E 2

„Ova arhitektura nadmašuje svoju prethodnu verziju u pogledu fotorealizma i sličnosti sa opisom. Pored postignuća u performansama, također nudi dodatne funkcije osim jednostavne sinteze teksta u sliku, kao što su:

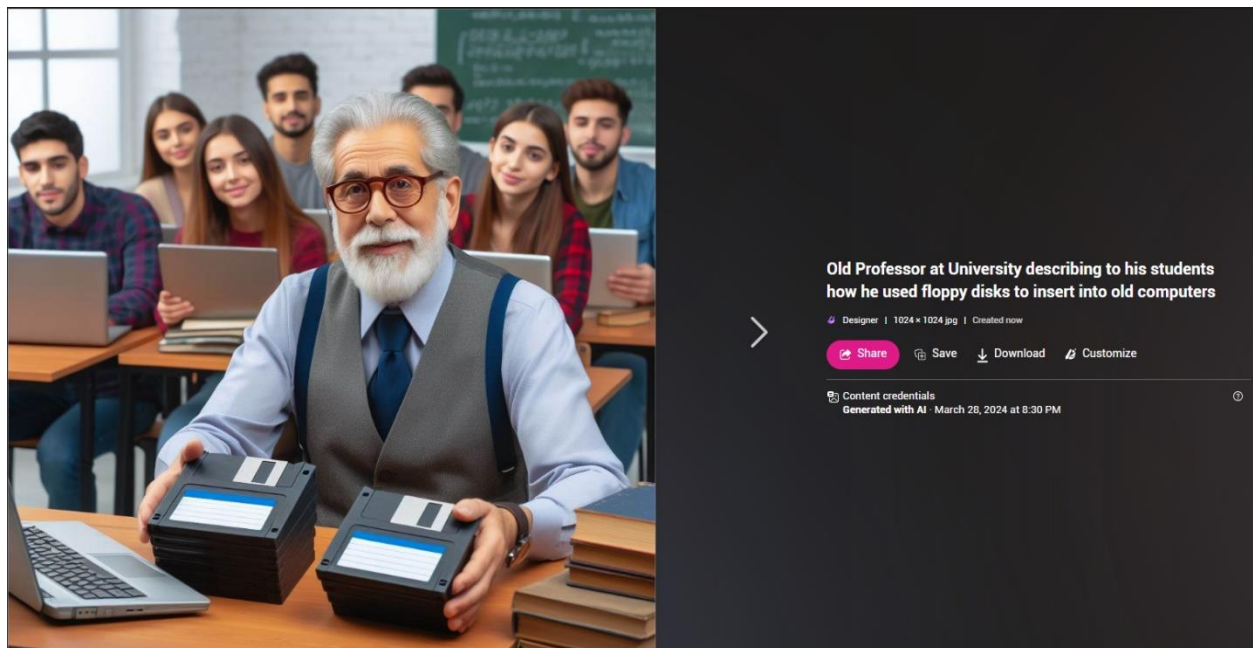
- Inpainting (generisanje određenih objekata i njihovo kombinovanje u veću sliku)
- Outpainting (kreiranje vizuelnih proširenja ulazne slike)
- Kreiranje varijacija ulaznih slika

Također je vrijedno napomenuti otvorenu implementaciju DALL-E-a nazvanu Craiyon (ranije poznatu kao DALL-E Mini, preimenovano radi prava kompanije OpenAI-a), napravljenu od strane nezavisnog programera i inženjera mašinskog učenja Boris Dayma. DALL-E Mini nije mogao dostići kvalitet originalne DALL-E arhitekture, ali je stekla određenu popularnost izvan AI zajednice putem internet mimova zbog svojih često smiješnih rezultata“ [3].



Slika 27 - Sadržaj slika generisan pomoću Craiyon softvera

“Pored ova dva velika projekta, prva polovina 2022. godine je vidjela pojavu mnogih drugih interesantnih projekata. U maju, Google je najavio svoj sistem generisanja slika Imagen, koji bi mogao da se takmiči sa DALL-E, a u junu je objavio Parti, koji je još jedan model teksta u sliku. Također su značajne publikacije MicroSoft-a o njihovom pristupu generisanja slika – NUWA ili NUWA Infinity. Svaka od ovih arhitektura je donijela zanimljive rezultate i varijacije na prethodno razvijene tehnike. Nažalost, nijedan od njih nije dostupan za testiranje široj javnosti i njihov detaljan opis je izvan našeg opsega“ [3].



Slika 28 - Slika generisana pomoću DALL-E 3 sa unosom "Stariji Profesor na Univerzitetu predstavlja svojim studentima kako je koristio Floppy Disk-ove"

ZAKLJUČAK

Analiziranjem ovog seminarskog rada, možemo zaključiti da viještačka inteligencija ima ogroman potencijal u transformaciji kompjuterske grafike. Od generisanja video sadržaja do stvaranja umjetničkih dijela, AI tehnologije otvaraju nove mogućnosti koje će oblikovati način na koji stvaramo digitalnu umjetnost.

Međutim, sa svim napretcima dolaze izazovi i pitanja koje treba razmotriti kao što su pitanja o etici, autentičnosti i sigurnosti ostaju važna, a napredak tehnologije zahtijeva odgovarajuće regulacije i smjernice kako bi se osiguralo da se viještačka inteligencija koristi na odgovoran i transparentan način.

Budućnost viještačke inteligencije u kompjuterskoj grafici će zavisiti od naše sposobnosti da balansiramo između inovacija i odgovornosti, a i da iskoristimo njen potencijal na način koji koristi svima.

POPIS SLIKA

Slika 1 - Vještački mozak	4
Slika 2 - Logo "Dartmouth" koledža	5
Slika 3 - Mašinsko učenje i njeni pod-tipovi	5
Slika 4 - Fotografija Američke vojske u korištenju ENIAC-a.....	6
Slika 5 - IBM 702 - Prvi računar korišten od strane prvih istraživača vještačke inteligencije.....	7
Slika 6 - Portret Alan Turing-a	7
Slika 7 - "Deep Blue" super-kompjuter	9
Slika 8 - Primjer stvarnog videa pretvoren u animaciju pomoću AI-a	10
Slika 9 - Prvi primjeri korištenja "Text-to-Video" tehnologije koja je malo uznemirujuća	11
Slika 10 - Primjer "Deep Fake" sadržaja	12
Slika 11 - AppleVision Pro	13
Slika 12 -- Tačka gledišta AppleVision Pro korisnika.....	13
Slika 13 - Struktura GAN mreže.....	14
Slika 14 - Generator i diskriminator u praksi.....	14
Slika 15 - Automobil koji ne postoji, kreiran pomoću GAN tehnologije	15
Slika 16 - Limitacije GAN tehnologije	16
Slika 17 - Logo OpenAI-a i logotip DALL-E softvera.....	17
Slika 18 - Arhitektura Transforvera u DALL-E	17
Slika 19 - Slika generisana od strane DALL-E koristeći input "fotelja u obliku avokada"	18
Slika 20 - "Putnik kroz vrijeme u gužvi" generisan od strane DeepDaze	19
Slika 21 - "Baloni iznad ruševina grada" generisano pomoću BigSleep-a	20
Slika 22 - Sadržaj generisan VQGAN-CLIP metodom	21
Slika 23 - Arhitektura CLIP difuzije.....	22
Slika 24 - Slika generisana pomoću CLIP Guided Diffusion-a na upit "Žena stoji u parku"	24
Slika 25 - "Mehanička ptica" kreirana četvrtom verzijom MidJourney softvera	25
Slika 26 - "Astronaut koji jaše konja u foto-realističnom stilu" generisano pomoću DALL-E 2.	26
Slika 27 - Sadržaj slika generisan pomoću Craiyon softvera	27
Slika 28 - Slika generisana pomoću DALL-E 3 sa unosom "Stariji Profesor na Univerzitetu predstavlja svojim studentima kako je koristio Floppy Disk-ove"	28

CITATNI IZVORI

- [1] Multiple, »Wikipedia,« Wikipedia, 28 3 2024. [Mrežno]. Available: https://en.wikipedia.org/wiki/History_of_artificial_intelligence. [Pokušaj pristupa 28 3 2024].
- [2] VideoZen, »Medium,« 29 3 2024. [Mrežno]. Available: <https://medium.com/@videozen/the-evolution-of-ai-in-video-generation-a-historical-perspective-8bc46d8bf242>. [Pokušaj pristupa 29 3 2024].
- [3] M. Chruściński, »sii,« 29 3 2024. [Mrežno]. Available: <https://sii.pl/blog/en/a-brief-history-of-ai-powered-image-genera>. [Pokušaj pristupa 29 3 2024].