

VAIZDŲ GENERAVIMAS NAUDOJANT DIRBTINIO INTELEKTO MODELĮ STABLE DIFFUSION

Emilis Raubickas, lekt. Gražina Tautvydienė

Šiaulių valstybinė kolegija, Aušros al. 40, Šiauliai

Anotacija. Dirbtinis intelektas tobulėja labai greitai, todėl reikia gilinti žinias apie šiuos modelius, išvelgti naujų technologijų naudą kasdieniniame gyvenime ir pritaikyti kuriant įvairias naujoves. Vaizdų generavimas naudojant dirbtinio intelekto modelius yra labai svarbi ir aktuali technologijos sritis. Vienas iš naujausių ir pažangiausių modelių šioje srityje yra *Stable Diffusion*.

Pagrindiniai žodžiai: dirbtinis intelektas, vaizdo generavimas, *stable diffusion*.

Įvadas

Stable Diffusion modelis leidžia generuoti vaizdus iš teksto įvesties naudojant dirbtinį intelektą (DI).

Šis modelis turi didelį potencialą įvairiose srityse. Pirmiausia, jis gali būti vertingas įrankis reklamos ir marketingo srityse. Tai leidžia greitai sukurti kokybiškus vaizdus, atitinkančius specifinius reikalavimus ir tikslus.

Be to, medicinos srityje teksto į vaizdą technologija gali būti labai naudinga. Galima generuoti diagnostinius vaizdus, pagrįstus pacientų simptomais arba tyrimų rezultatais. Tai gali pagerinti gydytojų gebėjimus diagnozuoti ligas ir padėti priimant sprendimus dėl gydymo.

Taip pat pramogų industrija gali pasinaudoti šia technologija. Vaizdų generavimas pagal teksto aprašymus gali būti naudojamas filmų ir žaidimų kūrimo. Tai padidina kūrėjų kūrybinį potencialą ir leidžia greitai vizualizuoti idėjas bei sukurti įspūdingus vaizdus.

Tyrimo objektas – dirbtinio intelekto vaizdų generavimo modelis *Stable Diffusion*

Tyrimo tikslas – išanalizuoti vaizdų generavimą naudojant dirbtinio intelekto modelį *Stable Diffusion*

Tyrimo uždaviniai:

1. Apibrėžti dirbtinio intelekto sampratą.
2. Išanalizuoti dirbtinio intelekto vaizdų generavimo modelį ir jo funkcijas.
3. Išsiaiškinti dirbtinio intelekto vaizdų generavimo modelio pritaikymo galimybes.

Tyrimo metodika – mokslinės literatūros ir dokumentinių šaltinių analizė, skaitmeninis eksperimentas.

Dirbtinio intelekto samprata

Dirbtinis intelektas (DI) – kompiuterių mokymosi ir sprendimų priėmimo technologijų rinkinys, kuris leidžia kompiuteriui atlikti užduotis, kurios reikalauja žmogaus intelekto. Tai yra mokslo ir inžinerijos sritis, kurią sudaro skirtingi kompiuterių algoritmai ir metodai, kuriuos galima naudoti mokantis iš duomenų, sprendžiant problemas, suprantant kalbą ir vaizdus, automatizuojant procesus ir kt. [8]

Vienas iš DI modelių pavyzdžių yra "*text to image*" (tekstas į vaizdą) modelis. Tai yra modelis, kuris gali paversti tekstinę informaciją į vizualią informaciją, pvz., paveikslus. Šie modeliai yra labai naudingi, nes jie gali padėti kurti vaizdinius, kai nėra galimybės juos fotografuoti arba sukurti piešiant ranka [7, 8]

Dirbtinio intelekto vaizdų generavimo modelis ir jo funkcijos

Stable Diffusion – dirbtinio intelekto modelis, kuris gali generuoti realistiškus vaizdus iš tekstinių aprašymų. Šis modelis sukurtas 2022 m. *Stable Diffusion* palaiko galimybę kurti naujus vaizdus nuo nulio naudojant tekstinę įvestį, kurioje aprašomi elementai, kurie turi būti įtraukti arba neįtraukti į išvestį. Esamus vaizdus dirbtinis intelektas gali sugeneruoti taip, kad į juos būtų įtraukti nauji elementai. Be to, modelis taip pat leidžia naudoti užuominas, kad iš dalies pakeistų esamus vaizdus.[6]

Stable Diffusion naudojimas *ONLINE*

Stable Diffusion 2.1 Demo yra pradinė šio modelio versija, kurią gali išbandyti kiekvienas vartotojas. Ši versija nereikalauja galingo kompiuterio ar gebėjimų perprasti ir koreguoti įvairius įvesties nustatymus, bet šios versijos išgaunami vaizdų rezultatai nėra visiškai tikslūs ir kokybiški (žr. 1 pav.) [8].

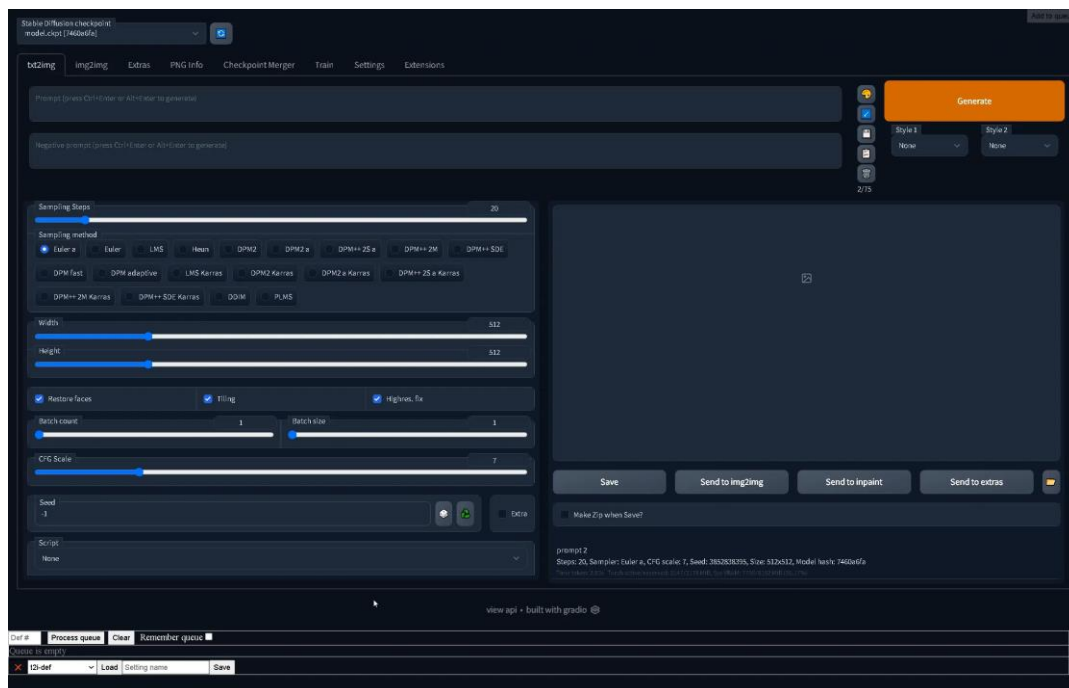


1 pav. *Stable Diffusion* versija *ONLINE* režimu

(Nuoroda į modelį– [Stable Diffusion 2-1 - a Hugging Face Space by stabilityai](#))

Įdiegiama versija AUTOMATIC1111 GUI (Grafinė vartotojo sąsaja)

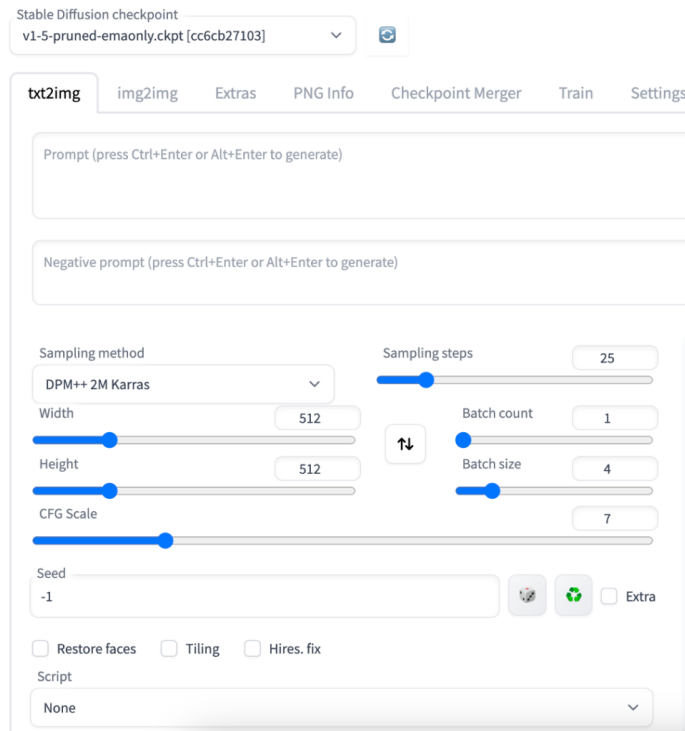
AUTOMATIC1111 (sutrumpintai A1111) yra skirta pažengusiems vartotojams. Įdiegiama versija yra nemokama ir šią sąsają galima naudoti per "Google Colab" platformą, taip pat parsisiųsti "Windows" arba "Mac". Minimalūs reikalavimai apima ne mažiau kaip 6 GB VRAM turintį grafinį procesorių, kuris yra daugumoje šiuolaikinių NVIDIA vaizdo plokščių. Be to, reikia maždaug 10 GB kietojo disko atminties ir suderintų kitų kompiuterio komponentų (žr. 2 pav.) [10].



2 pav. Grafinė vartotojo sąsaja

Grafinėje vartotojo sąsajoje pakeisti modelio funkciniai nustatymai: (žr. 3 pav.) [1,4]

- Teksto įvestis (*Prompt*): rašomas tekstas ką norima matyti vaizde. Tekstas formuojamas konkrečiai ir išsamiai, naudojant įvairius raktažodžius.
- Neigiama įvestis (*Negative Prompt*): rašomas tekstas, ko nenorima matyti. Dažniausiai naudojamos universalios neigiamos įvestys.
- Bandymų metodas (*Sampling method*): vaizdo generavimo greičio ir kokybės gerinimui buvo naudojamas DPM++ 2M Karras atrinkimo proceso algoritmas.
- Bandymų žingsniai (*Sampling steps*): daugeliu atvejų tinka 25 žingsniai, bet geresnei vaizdo kokybei gauti žingsnių skaičius didinamas.
- Plotis ir aukštis: išvesties vaizdo dydis. Pavyzdžiui, portretiniam vaizdui, kurio kraštinių santykis 2:3, nustatomas plotis 512, o aukštis - 768 pikselių.
- CFG skalė: parametras, kuriuo galima valdyti, kiek modelis turėtų atspindėti įvestyje (*Prompt*) pateiktą tekstą.



3 pav. Grafinės vartotojo sąsajos nustatymai

Vaizdo generatoriui pateikiama teksto įvestis turi būti išsami ir konkreti. Taip pat įvedamas tekstas turi būti rašomas anglų kalba. Teksto įvestį galima kurti pagal raktažodžių kategorijas, pradėti nuo objekto iki apšvietimo vaizde. Galima papildyti įvestį žodžiais ir stebėti, kaip rezultatas skiriasi bei jį lyginti su prieš tai buvusiais vaizdais. [2,3,9]

Įvesties raktažodžių kategorijos yra šios: [2,3,9]

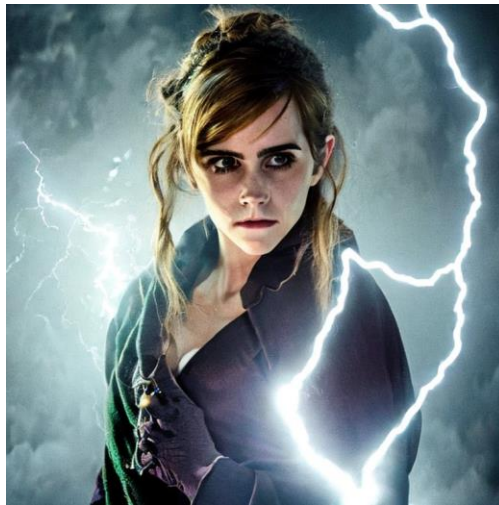
- Tema (*Subject*)
- Medija (*Media*)
- Stilius (*Style*)
- Atlikėjas (*Artist*)
- Interneto svetainė (*Website*)
- Skiriamoji geba (*Resolution*)
- Papildoma informacija (*Additional details*)
- Spalva (*Colour*)
- Apšvietimas (*Lightning*)

DI vaizdo generatoriaus funkcinėms galimybėms apžvelgti buvo atliktas skaitmeninis eksperimentas, kuriame numatyta pateikti tam tikrą trumpą aprašymą ir peržiūrėti, kaip modelis sugeneruoja paveikslėlius pagal aprašymą. Galima pabandyti naudoti skirtingus aprašymus ir peržiūrėti, kaip dirbtinis intelektas juos interpretuoja ir sugeneruoja vizualius atvaizdus. Tai gali padėti geriau suprasti, kaip veikia šis modelis ir kokie yra jo privalumai ir trūkumai.

- Vaizdo generavimas – objektas (*Subject*)

Objektas yra tai, ką norima matyti nuotraukoje. Dažna klaida - nepakankamai aprašyti objektai. Kuriama Emma Watson, kaip burtininkės išvaizda. Tarkime ji yra galinga, paslaptina ir naudoja žaibų magiją. Norima, kad ji atrodytų idomiai su detaliai aprašyta apranga (žr. 4 pav.).

Įvestis (*Prompt*) – Emma Watson as a powerful mysterious sorceress, casting lightning magic, detailed clothing (Emma Watson kaip galinga paslaptina burtininkė, skleidžianti žaibų magiją, detalūs drabužiai).



4 pav. Eksperimento rezultatas Nr. 1

- Vaizdo generavimas: **stilius** (*Style*)
Tai reiškia meninį vaizdo stilių. Pavyzdžiui, impresionistinis, siurrealistinis ir kt. Prie tekstinės įvesties nurodomi stiliai: hiperrealistinis, fantastinis, siurrealistinis, pilnas kūnas (žr. 5 pav.).

Įvestis (*Prompt*) – Emma Watson as a powerful mysterious sorceress, casting lightning magic, detailed clothing, digital painting, hyperrealistic, fantasy, Surrealist, full body. (Emma Watson kaip galinga paslaptina burtininkė, metanti žaibų magiją, detalūs drabužiai, skaitmeninė tapyba, hiperrealistinė, fantazija, siurrealistinė, pilnas kūnas).



5 pav. Eksperimento rezultatas Nr. 2

- Vaizdo generavimas: neigiama įvestis (*Negative prompt*)

Neigiamos įvestys naudojamos panaikinti įvairias detales, modifikuojant gaunamą rezultatą, ryškinant vaizdą ir padarant jį realistiškesniu. Neigiamos įvesties žodžių papildymas gali būti kartotinio proceso dalis. Raktažodžiai gali būti objektai arba kūno dalys, kurių norima vengti (žr. 6 pav.).

Naudojama universali neigiama tekstinė įvestis: *ugly, tiling, poorly drawn hands, poorly drawn feet, poorly drawn face, out of frame, extra limbs, disfigured, deformed, body out of frame, bad anatomy, watermark, signature, cut off, low contrast, underexposed, overexposed, bad art, beginner, amateur, distorted face* (negražus, plytelės, blogai nupieštos rankos, blogai nupieštos kojos, blogai nupieštas veidas, iškritęs iš kadro, papildomos galūnės, išsigimęs, deformuotas)

Įvestis (*Prompt*) – *Emma Watson as a powerful mysterious sorceress, casting lightning magic, detailed clothing, digital painting, hyperrealistic, fantasy, Surrealist, full body, by Stanley Artgerm Lau and Alphonse Mucha*



6 pav. Eksperimento rezultatas Nr. 3

Galima pastebėti kaip rezultatas skiriasi į įvestį pridėjus keletą raktinių žodžių. Kuriant su *Stable Diffusion* gerus vaizdus, rezultatus galima išgauti ir nepanaudojus visų raktinių žodžių kategorijų, bet nuolat tobulinant įvestį ir eksperimentuojant.

Dirbtinio intelekto vaizdų generavimo modelio pritaikymo galimybės [5]

DI vaizdų generavimo modelis *Stable Diffusion* gali būti pritaikomas:

- produktų katalogų vaizdų generavime – gali būti naudojama katalogams skirtiems produktų atvaizdams generuoti, todėl nebereikia rankiniu būdu fotografuoti produktų;
- logotipų kūrimo – galima naudoti įmonių logotipams kurti, taip sutaupant laiko ir pinigų dizaino išlaidoms ir kt.;
- vaizdų redagavime – galima naudoti automatiniam vaizdų redagavimui, todėl įmonės gali greitai ir lengvai keisti esamus vaizdus;
- vaizdų retušavime – galima naudoti automatiniam vaizdų retušavimui, todėl įmonės gali greitai ir lengvai pagerinti vaizdų kokybę.

Išvados

- *Stable Diffusion* yra vienas iš dirbtinio intelekto modelių, kuris gali būti naudojamas vaizdų generavimui iš tekstinių aprašymų.
- *Stable Diffusion* yra atvirojo kodo modelis, tai reiškia, kad kiekvienas, turintis reikiamų techninių įgūdžių, gali jį įsidiegti ir panaudoti fantastinių ir realistinių vaizdų generavimui.
- Išanalizavus *Stable Diffusion* modelį ir jo funkcijas, numatytos modelio pritaikymo galimybės.
- Modeliui tobulėjant, skaitmeninių eksperimentų būdu, įdomu stebėti jo pažangą ateityje.

Informacijos šaltiniai

1. Artificial intelligence (AI). Prieiga per internetą 2023-05-07: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>
 2. Binxu Wang, Understanding Stable Diffusion from "Scratch". Prieiga per internetą 2023-05-07: <https://scholar.harvard.edu/binxuw/classes/machine-learning-scratch/materials/stable-diffusion-scratch>
 3. Dehouche N, Dehouche K, What's in a Text-to-Image Prompt? The Potential of Stable Diffusion in Visual Arts Education, 2023. Prieiga per internetą 2023-05-09: <https://arxiv.org/ftp/arxiv/papers/2301/2301.01902.pdf>
 4. Diab M. ir kt., Stable Diffusion Prompt Book, 2022. Prieiga per internetą 2023-05-09: <https://cdn.openart.ai/assets/Stable%20Diffusion%20Prompt%20Book%20From%20OpenArt%2010-28.pdf>
 5. Rost M., Andreasson S., Stable Walk: An interactive environment for exploring Stable Diffusion outputs, 2023. Prieiga per internetą 2023-05-09: <https://ceur-ws.org/Vol-3359/paper10.pdf>
 6. SD Compendium. Prieiga per internetą 2023-05-06: <https://www.sdcompendium.com/doku.php?id=start>
 7. Stable diffusion links: news, links, and more. Prieiga per internetą 2023-05-07: <https://reentry.org/niakonichan>
 8. Stable Diffusion Online. Prieiga per internetą 2023-05-20: <https://stablediffusionweb.com/>
 9. Stable Diffusion prompt: a definitive guide. Prieiga per internetą 2023-05-07: <https://stable-diffusion-art.com/prompt-guide>
- Stable Diffusion WebUI AUTOMATIC1111: A Beginner's Guide. Prieiga per internetą 2023-05-07: <https://stable-diffusion-art.com/automatic1111>