Application of text-to-image translation algorithms in medicine: A systematic review

Michelle Mai, BS,[a] Fatima N. Mirza, MD,[a] and Christopher DiMarco, MD[a,b]

Medicine is rapidly evolving with the integration of advanced technologies, particularly in response to user-prompted artificial intelligence. The development of artificial intelligence algorithms that utilize image generation poses many benefits in the realm of diagnostics, patient care, domain knowledge, and medical education. Such advanced technologies include text-guided image manipulation algorithms like generative adversarial networks, transformer-based models, and zero-shot transfer learning through OpenAI's DALL-E2. This systematic review delves into these recent developments of text-guided image manipulation algorithms and applications of these algorithms in medicine. Current research has utilized image generation algorithms in denoising, image modality transfer, segmentation, and data augmentation. However, the incorporation of text-guided inputs for image generation and the integration of these algorithms into medical applications has been less studied. Text-to-image algorithms in medicine have potential in medical education initiatives, increasing data scarcity for rare diseases and improving data quality. However, given their recent developments, text-to-image algorithms may also perpetuate biases against minorities and present limited domain knowledge. (JAAD Reviews 2024;2:88-96.)

Key words: artificial intelligence; computer vision; dermatology; lesion detection; medical education; text-to-image.

INTRODUCTION

Medicine is rapidly changing, with the advent of adaptable technologies like artificial intelligence (AI). Machines are now able to make more accurate diagnoses based on image cues, warn physicians when mistakes are entered on electronic health records, and use generative AI models to pass board examinations for several specialties.[1-4] Incorporating these elements of technology into a physician's workflow may facilitate better quality care while streamlining administrative tasks.[5] Furthermore, generative AI may change current medical education practices given its ability to filter numerous resources that are publicly available. From the patient's perspective, the usage of trained algorithms to make decisions on patient care may confer increased accuracy and aid in improving health outcomes.[6] The study of generative AI models, particularly the recent development of generative image models, is necessary to understand the profound impacts the widespread utilization of these algorithms may have on health care and the medical field at large.[7] Generative adversarial networks (GANs), modified GANs for text-to-image extrapolation, transformer-based algorithms, and Open-AI's DALL-E 2 and contrastive language-image pretraining (CLIP) models are the state-of-the-art methodologies for text-to-image translation. This systematic review explores the technical algorithms developed and discusses the utilization of text-to-image algorithms in medicine.

METHODS

This systematic review (PROSPERO CRD42024501235) was conducted following guidelines from the Preferred Reporting Items for Systematic Review and Meta-Analysis. Studies selected for this systematic review included

From the Department of Dermatology, Warren Alpert Medical School, Providence, Rhode Island[a]; and Department of Pathology and Laboratory Medicine, Warren Alpert Medical School, Providence, Rhode Island.[b]

Funding sources: None.

IRB approval status: Not applicable.

Accepted for publication September 9, 2024.

Correspondence to: Christopher DiMarco, MD, Department of Dermatology, Warren Alpert Medical School, 593 Eddy St., Providence, RI 02903. E-mail: chris.dimarco@brownphysicians.org.

2950-1989

https://doi.org/10.1016/j.jdrv.2024.09.003

full-text, English-language journal articles published from January 2019 through January 2024. All original research studies discussing any algorithm choice in which text was used to guide, generate, or edit images were included. Studies and reviews discussing implementations within medical specialties were included. Studies that did not explicitly mention text-to-image algorithms were not considered. Three databases including Pubmed, EMBASE, and IEEE were searched for "generative AI" OR "image manipulation" OR "image generator" OR "text to image" OR "stable diffusion." The results from this search were coupled with "dermatology" and subsequently with "medicine" to provide dermatology and general medicine applications. The data collected from these studies included year of publication, AI model type, model application in medicine, and medical specialty. To minimize bias errors, study selection and data extraction was conducted independently by reviewers. A narrative synthesis of the data was collected into an Excel sheet for summary analysis through simplified text and tables.

CAPSULE SUMMARY

- Various algorithms relevant to text-to-image translation are at the forefront of artificial intelligence research.
- These algorithms offer promising opportunities to enhance medical education, improve data quality, and facilitate diagnosis and patient care, while also highlighting the need for further research to address challenges.

RESULTS

A total of 3095 articles were retrieved from Pubmed, IEEE, and Embase. Two thousand, six hundred twenty-three articles were removed because of duplication. An additional 267 articles were removed to ensure original research articles were written in the English language. Twenty-three records were nonaccessible full-text articles. Thorough title and abstract perusal for topic alignment in the inclusion criteria for medical applications of text-to-image synthesis algorithms and medical applications yielded a total of 38 studies, 16 of which were development of new generators and 22 of which were analysis of text-to-image algorithms in a medical context (Fig 1).

Of the 16 studies of new algorithms (Table I),[8-23] 9 studies include variants of GANs and 5 more recent studies utilize diffusion models based on denoising diffusion probabilistic models and denoising diffusion implicit models. A common theme among these works is that they encode text inputs into lower dimensional space to inform image generation or editing. In 2021, Radford et al[24] introduced CLIP, a large pretrained model that correlates texts' lower dimensional space with that of images. CLIP has since been adopted widely in text-to-image works. Table I[8-23] also shows that diffusion models have become more popular than GANs in recent years. Compared to GANs, diffusion models are more stable to train, yield higher-quality results, and offer greater controllability.

Of the 22 included studies of text-to-image medical appli cations, 14 studies involved OpenAI's DALL-E text-to-ima ge generator (Table II).[25-46] Other popular methods include Midjourney, GLIDE, and variants of GANs as descr ibed above. All the studies had novel applications including medical education through developing images of rare diseases for medical trainees, assessing diverse representations and perpetuation of societal biases via AI algorithms, and analyzing the landscape of knowledge an AI generator may have in a specialized field. Six papers evaluated a text-to-image generator's ability to synthesize images for a particular use case realistically and accurately. Six papers also evaluated text-to-image algorithms for use in medical and patient education. Five papers describe text-to-image generator's roles in perpetuating biases in medicine such as ageism, genderism, and racism. Five papers describe the involvement of text-to-image generators in improving data quality through increasing data scarcity of rare disease processes and denoising images. Studied specialties span ophthalmology, neuro-ophthalmology, dermatology, plastic surgery, general surgery, radiology, cardiology, oncology, geriatrics, pediatrics, and medicine.

DISCUSSION

Algorithms for text-to-image generation

GANs are a class of popular machine learning algorithms developed and modified for use in the field of computer vision. A GAN involves 2 neural networks: a generator, which creates synthetic images, and a discriminator, which evaluates the realism of these images.[47] The GAN utilizes this dynamic feedback system to produce increasingly realistic images.[48]

GANs have also been modified for text-guided image synthesis. Text prompts like "create a brown mole with irregular borders" are transformed into numeric representations called embeddings that direct the generation of the specific image.[8,49] These embeddings serve as an additional input for

Abbreviations used:

AI: artificial intelligence
CLIP: contrastive language-image pretraining
GAN: generative adversarial network
GPT: generative pretrained transformer

the generator to make more focused images. Many free-form text-to-image GANs exist that follow this general framework (Fig 2), despite potential variations in the methodology transforming the free text into embeddings and in the specific GAN architecture (Table I).[8-23,50-52]

Although GANs are useful, they may struggle with generating images from complex text descriptions.[53] Transformer-based algorithms resolve this using self-attention, which enables them to understand relational features of specific words or phrases (ie, the relationship between a subject and verb may be weighed more heavily than the relationship between a subject and a preposition).[20,54] These models can generate images by integrating this contextual understanding. Commonly known transformer-based

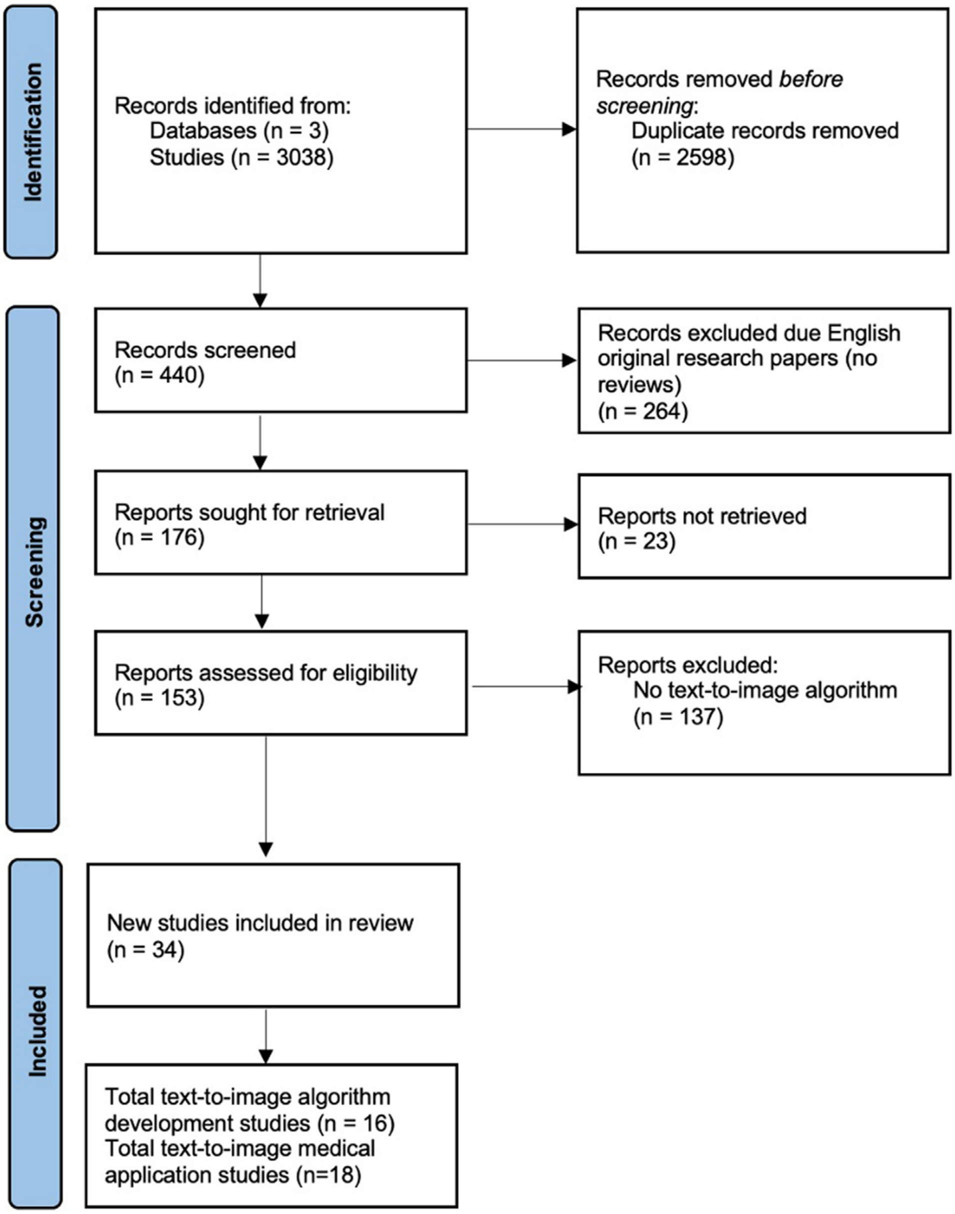

Fig 1. PRISMA flowchart for articles for review. *PRISMA*, Preferred Reporting Items for Systematic Review and Meta-Analysis.

Table I. Text-to-image algorithms developed from 2019 to 2023

Title	Authors	Model
Generative adversarial text to image synthesis	Reed et al[8]	GANs
StackGAN++: realistic image synthesis with stacked generative adversarial networks	Zhang et al[9]	Stack-GAN++
Latent Dirichlet allocation-based generative adversarial networks	Pan et al[10]	LDAGAN
KT-GAN: knowledge-transfer generative adversarial network for text-to-image synthesis	Tan et al[11]	KT-GAN
Multisentence auxiliary adversarial networks for fine-grained text-to-image synthesis	Yang et al[12]	MA-GAN
SAM-GAN: self-attention supporting multistage generative adversarial networks for text-to-image synthesis	Peng et al[13]	SAM-GAN
Zero-shot text-to-image generation	Ramesh et al[14]	CLIP
DR-GAN: distribution regularization for text-to-image generation	Tan et al[15]	DR-GAN
Text-guided human image manipulation via image-text shared space	Xu et al[16]	AdaIN decoder and GAN generator
Vector quantized diffusion model for text-to-image synthesis	Gu et al[17]	VQ-Diffusion, DDPM
DiffusionCLIP: text-guided diffusion models for robust image manipulation	Kim et al[18]	DiffusionCLIP
Blended diffusion for text-driven editing of natural images	Avrahami et al[19]	CLIP, DDPM
Muse: text-to-image generation via masked generative transformers	Chang et al[20]	Masked generative transformers
Word self-update contrastive adversarial networks for text-to-image synthesis	Xiao et al[21]	WSC-GAN
SINE: SINgle image editing with text-to-image diffusion models	Zhang et al[22]	DDPM
Plug-and-play diffusion features for text-driven image-to-image translation	Tumanyan et al[23]	DDPM, DDIM

CLIP, Contrastive language-image pretraining; *DDIM*, denoising diffusion implicit model; *DDPM*, denoising diffusion probabilistic model; *DR*, distribution regularization; *GAN*, generative adversarial network; *KT*, knowledge transfer; *LDAGAN*, latent dirichlet allocation-based generative adversarial network; *MA*, multisentence auxiliary; *SAM*, self-attention supporting multistage; *WSC*, word self-update contrastive.

large language models such as generative pretrained transformer (GPT), bidirectional encoder representations from transformers, and Muse have been pretrained using a corpus collection of publicly available text data.

More recently, zero-shot learning algorithms have been used in text-to-image generation. These algorithms enhance photo-realism with limited data.[14,45,55] In 2022, OpenAI released DALL-E 2 which employs a zero-shot learning algorithm called CLIP (Fig 3).[24] CLIP has been integrated into numerous generation algorithms caused by its efficiency.

Similarly, diffusion models have emerged as a popular method for text-to-image generation and editing tasks, offering a novel way to create high-quality and semantically accurate images from text prompts. Diffusion models initially generate random noise and gradually transform it into a coherent image that matches a given textual prompt.[17] Diffusion models utilize U-Net, which was first developed for biomedical imaging and helps preserve image details step-by-step.[56] CLIP and diffusion models can be used together for the most efficient and realistic image generation.[18,19]

Application of text-to-image algorithms in medicine

The utilization of text-to-image synthesis algorithms is relatively new—even outside of medical applications. Research on the development of new algorithms such as stable diffusion models and zero-shot techniques have led to increased prominence of text-to-image synthesis. Currently, these algorithms utilize random image data sets for comprehensive and generalizable image generation. Therefore, the lack of specificity in medical imaging has resulted in limited applications in medicine and dermatology.

Despite the limited study in the medical realm, initial research on medical applications of text-to-image synthesis is promising. Images pertinent to a variety of fields such as cardiology, dermatology, ophthalmology, and surgery have been identified as use cases for image generation.[57] The majority of these studies use a publicly available image-generative AI like DALL-E or Midjourney, which increases accessibility and usage by all users or clinicians regardless of technical ability. Implementation of text-to-image generation can be used for denoising and reconstructing images,

Table II. Medical applications of text-to-image generators

Author	Model	Usage	Medical specialty
Wan et al[25]	GANs	Domain knowledge	Ophthalmology
Ko et al[26]	GANs	Data scarcity	Dermatology
Kather et al[27]	GLIDE (diffusion model)	Domain knowledge	Oncology
Williams et al[28]	DALL-E 2	Domain knowledge	Cardiology
Shavlokhova et al[29]	GLIDE (diffusion model)	Data quality	Dermatology
Cheraghlou, 2023[30]	DALL-E 2	Domain knowledge	Dermatology
Noel[31]	DALL-E, Stable Diffusion, and Craiyon v3	Education	Medicine
Manocha et al[32]	DALL-E 2	Diversity/representation	Geriatrics
Waisberg et al[33]	DALL-E 2	Education	Neuro-ophthalmology
Sanchez et al[34]	DALL-E 2	Diversity/representation	Ophthalmology
Waikel et al[35]	DALL-E	Education	Pediatrics
Kenig et al[36]	Craiyon, DALL-E, and Midjourney	Diversity/representation	Plastic surgery
Koljonen[37]	DALL-E 2	Education	Plastic surgery
Lim et al[38]	DALL-E 2, Midjourney, BlueWillow, ChatGPT 4, and Google's Bard	Education	Plastic surgery
Khader et al[39]	DALL-E 2, Imagen, and Stable Diffusion	Data quality	Radiology
Adams et al[40]	DALL-E 2	Domain knowledge	Radiology
Ali et al[41]	DALL-E 2, Midjourney, and Stable Diffusion	Diversity/representation	Surgery
Cho et al[42]	GANs	Data scarcity	Dermatology
DeGrave et al[43]	GANs	Data scarcity	Dermatology
Zhu et al[44]	DALL-E 3	Education	Cardiology
Javan and Mostaghni[45]	Midjourney	Domain knowledge	Radiology
Ali et al[46]	DALL-E 2, Midjourney, and Stable Diffusion	Diversity/representation	Surgery

transferring image modalities (ie, manipulating a dermatoscopic representation of the skin into a digital image), image segmentation, and data augmentation.[58-60] Text-to-image generation in medicine, particularly in dermatology, has implications for advancing medical education, patient education, providing a source of ethical or deidentified data, and rare disease domain imaging. Medical education both for trainees and patients has been studied, finding that generative AI for 3D anatomy visualization or teaching difficult processes like 12-lead electrocardiograms are effective.[31,44] In dermatology, developing counterfactual images of melanoma helped evaluate undesirable features such as lesion pigmentation, even on background skin texture and color balance.[43] Patient education initiatives have been less studied; however, the impact of generative AI on patient education has been demonstrated through the use of ChatGPT and other software in lowering the reading level of educational texts.[61] Likewise, the usage of easy-to-interpret images may improve the understanding of certain educational topics. Generative imaging is important for developing comprehensive data sets important for representation and standardization. For example, Cho et al[42] developed a melanoma and nevus data set using unstandardized photos from the internet created by generative algorithms, showing the substantial role text-to-image synthesis has on providing quality, standardized, and deidentified data. Another application of generative AI in medicine involves rare disease domain imaging. Few studies have been done, especially because most AI algorithms may not be able to train on rare disease images and instead may rely on inference. However, a 2024 study by Waikel et al[62] found an application of generated images of Kabuki syndrome and Noonan syndrome to be noninferior to real images in helping pediatric residents recognize these syndromes. For dermatology, studies on rare disease imaging for dermatological conditions like different genodermatoses may also benefit from text-to-image synthesis algorithms.

Given the exciting potential for text-to-image synthesis, several studies emphasize several cautions with the dependency on such algorithms. Augmenting or generating new images involves synthesis of trained data; when specified to a medical domain, underrepresentation and underlying assumptions of specific racial and gender groups may only be exacerbated as suggested by Ali et al[46] and Manocha et al.[32] These studies found that generative AI algorithms propagated societal biases and were more likely to suggest that demographic representation of surgeons were the majority (98%) white

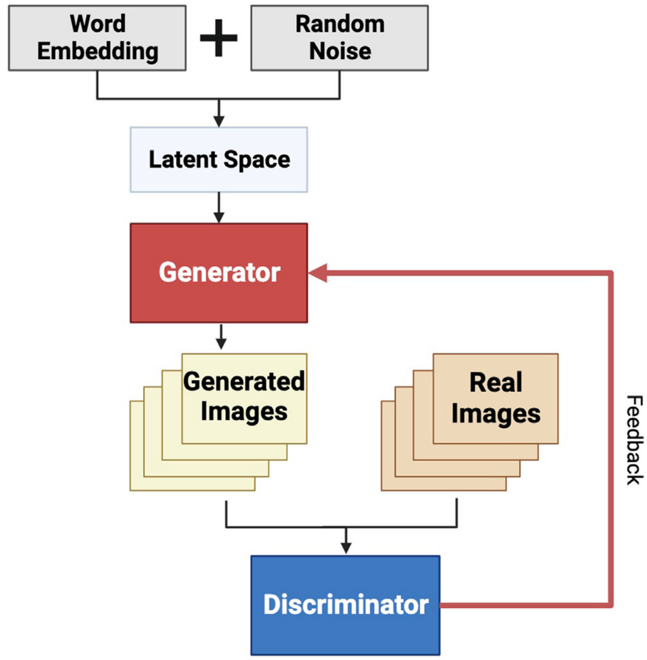

Fig 2. GAN-based framework for text-to-image synthesis. The latent space is a theoretical area that serves as a hub where the text embedding can be combined with random noise to better inform image generation. *GAN*, Generative adversarial network.

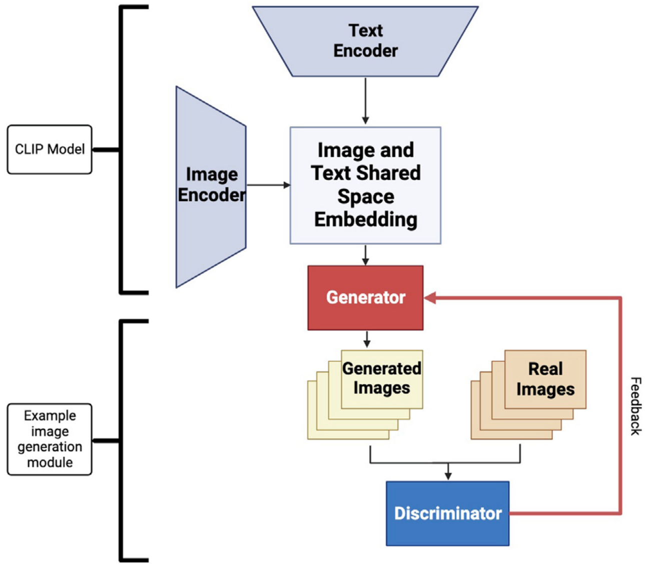

Fig 3. High-level view of CLIP mechanism and application to image generation module, in this case showing GAN architecture. *CLIP*, Contrastive language-image pretraining; *GAN*, generative adversarial network.

males. In fields such as plastic surgery or cosmetic dermatology, generative AI propagated directed standards of human beauty, affecting the protocol or procedures completed.[36] In addition, current user-facing text-to-image generators like DALL-E and Midjourney face limitations given their primary functionality in providing pseudorealistic, digital art renditions of life-like images.[38] Given the recency of image-generative AI, especially in medicine, further work in training generative AI algorithms specific to medicine may be necessary. To address the issue of perpetuating biases, studies on increasing and maintaining representation of all groups are important. Possible avenues of research could involve generalizability or representative data training sets. Furthermore, improving imaging data is key to enhancing AI research in the medical field. Developing robust dermatologic data sets using standardized imaging techniques with multiple modalities like dermatoscopy and digital cameras may be the first step to improving the current state of text-to-image generation.

The current state of text-to-image generation technology is promising with the increased usage of zero-shot learning and diffusion models to improve the efficiency and realism of prompted images. Advancements driven both by industry and academia have been made to increase accessibility through mobile access and to optimize functionality (memory storage and training-free learning).[63,64] Before efforts for clinical integration are made, algorithm adjustments to address baseline issues such as reliability and algorithm bias are needed.

CONCLUSIONS

Medicine is experiencing a rapid transformation through the integration of advanced technology and AI. This paper explores the emerging trend of text-guided image manipulation techniques and their potential applications in medicine. Various algorithms relevant to text-to-image translation, such as GANs, transformer-based text-to-image algorithms, zero-shot transfer learning text-to-image generation, and stable diffusion, are at the forefront of AI research and offer exciting possibilities for data augmentation in medicine. The integration of text-guided image manipulation techniques into medicine has the potential to unlock new tools for diagnosis, patient care and education, and medical education. While exciting, further research and development in text-to-image generation specific to a medical specialty or domain is needed. Overall, these text-to-image algorithms applied to medicine present a promising frontier in medical AI.

Thank you to Sean Zhan (PhD Candidate at the Massachusetts Institute of Technology) for his guidance and expertise on text-to-image models.

Conflicts of interest

None disclosed.

REFERENCES

1. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. *Radiology*. 2023;307(5):e230582. https://doi.org/10.1148/radiol.230582
2. Rau A, Rau S, Zoeller D, et al. A context-based chatbot surpasses trained radiologists and generic ChatGPT in following the ACR appropriateness guidelines. *Radiology*. 2023;308(1):e230970. https://doi.org/10.1148/radiol.230970
3. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887. https://doi.org/10.3390/healthcare11060887
4. Mirza FN, Lim RK, Yumeen S, et al. Performance of three large language models on dermatology board examinations. *J Invest Dermatol*. 2024;144(2):398-400. https://doi.org/10.1016/j.jid.2023.06.208
5. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. https://doi.org/10.1038/s41591-021-01614-0
6. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395(10236):1579-1586. https://doi.org/10.1016/S0140-6736(20)30226-9
7. Arora A, Arora A. Generative adversarial networks and synthetic patient data: current challenges and future perspectives. *Future Healthc J*. 2022;9(2):190-193. https://doi.org/10.7861/fhj.2022-0013
8. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative adversarial text to image synthesis. Paper presented at: Proceedings of the 33rd International Conference on Machine Learning; Proceedings of Machine Learning Research, 2016. https://proceedings.mlr.press/v48/reed16.html
9. Zhang H, Xu T, Li H, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks. *IEEE Trans Pattern Anal Mach Intell*. 2019;41(8):1947-1962. https://doi.org/10.1109/tpami.2018.2856256
10. Pan L, Cheng S, Liu J, et al. Latent Dirichlet allocation based generative adversarial networks. *Neural Netw*. 2020;132:461-476. https://doi.org/10.1016/j.neunet.2020.08.012
11. Tan H, Liu X, Liu M, Yin B, Li X. KT-GAN: knowledge-transfer generative adversarial network for text-to-image synthesis. *IEEE Trans Image Process*. 2021;30:1275-1290. https://doi.org/10.1109/tip.2020.3026728
12. Yang Y, Wang L, Xie Deng C, Tao D. Multi-sentence auxiliary adversarial networks for fine-grained text-to-image synthesis. *IEEE Trans Image Process*. 2021;30:2798-2809. https://doi.org/10.1109/TIP.2021.3055062
13. Peng D, Yang W, Liu C, Lü S. SAM-GAN: self-attention supporting multi-stage generative adversarial networks for text-to-image synthesis. *Neural Netw*. 2021;138:57-67. https://doi.org/10.1016/j.neunet.2021.01.023
14. Ramesh A, Pavlov M, Goh G, et al. Zero-shot text-to-image generation. Paper presented at: Proceedings of the 38th International Conference on Machine Learning; Proceedings of Machine Learning Research, 2021. https://proceedings.mlr.press/v139/ramesh21a.html
15. Tan H, Liu X, Yin B, Li X. DR-GAN: distribution regularization for text-to-image generation. *IEEE Trans Neural Netw Learn Syst*. 2023;34:10309-10323. https://doi.org/10.1109/tnnls.2022.3165573
16. Xu X, Chen YC, Tao X, Jia J. Text-guided human image manipulation via image-text shared space. *IEEE Trans Pattern Anal Mach Intell*. 2022;44(10):6486-6500. https://doi.org/10.1109/tpami.2021.3085339
17. Gu S, Chen D, Bao J, et al. Vector quantized diffusion model for text-to-image synthesis. 2022:10696-10706.
18. Kim G, Kwon T, Ye JC. DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation. Published online 2021. https://doi.org/10.48550/ARXIV.2110.02711
19. Avrahami O, Lischinski D, Fried O. Blended diffusion for text-driven editing of natural images. Published online 2021. https://doi.org/10.48550/ARXIV.2111.14818

20. Chang H, Zhang H, Barber J, et al. Muse: text-to-image generation via masked generative transformers. Published online 2023. https://doi.org/10.48550/ARXIV.2301.00704
21. Xiao J, Sun Y, Bi X. Word self-update contrastive adversarial networks for text-to-image synthesis. *Neural Netw*. 2023;167:433-444. https://doi.org/10.1016/j.neunet.2023.08.038
22. Zhang Z, Han L, Ghosh A, Metaxas DN, Ren J. Sine: single image editing with text-to-image diffusion models. 2023:6027-6037.
23. Tumanyan N, Geyer M, Bagon S, Dekel T. Plug-and-play diffusion features for text-driven image-to-image translation. Published online 2022. https://doi.org/10.48550/ARXIV.2211.12572
24. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. PMLR. Published online 2021. https://doi.org/10.48550/ARXIV.2103.00020
25. Wan C, Zhou P, Wu L, Wu Y, Shen J, Ye H. Diabetic retinopathy fundus image generation based on generative adversarial networks. *Chin J Exp Ophthalmol*. 2019;37(8):613-618. https://doi.org/10.3760/cma.j.issn.2095-0160.2019.08.005
26. Ko M, Kim D, Kim K. Accurate depth estimation of skin surface using a light-field camera toward dynamic haptic palpation. *Skin Res Technol*. 2019;25(4):469-481. https://doi.org/10.1111/srt.12675
27. Kather JN, Ghaffari Laleh N, Foersch S, Truhn D. Medical domain knowledge in domain-agnostic generative AI. *NPJ Digit Med*. 2022;5(1):90. https://doi.org/10.1038/s41746-022-00634-5
28. Williams MC, Williams SE, Newby DE. Artificial intelligence-based text-to-image generation of cardiac CT. *Radiol Cardiothorac Imaging*. 2023;5(2):e220297. https://doi.org/10.1148/ryct.220297
29. Shavlokhova V, Vollmer A, Zouboulis CC, et al. Finetuning of GLIDE stable diffusion model for AI-based text-conditional image synthesis of dermoscopic images. *Front Med (Lausanne)*. 2023;10:1231436. https://doi.org/10.3389/fmed.2023.1231436
30. Cheraghlou S. Evaluating dermatologic domain knowledge in DALL-E 2 and potential applications for dermatology-specific algorithms. *Int J Dermatol*. 2023;62(10):e521-e523. https://doi.org/10.1111/ijd.16683
31. Noel G. Evaluating AI-powered text-to-image generators for anatomical illustration: a comparative study. *Anat Sci Educ*. 2024;17(5):979-983. https://doi.org/10.1002/ase.2336
32. Manocha GD, Derenne N, Van Eck R, Jurivich D. Artificial intelligence perpetuates ageism via imagery selection. *J Am Geriatr Soc*. 2023;71:S56. https://doi.org/10.1111/jgs.18336
33. Waisberg E, Ong J, Masalkhi M, et al. Text-to-image artificial intelligence to aid clinicians in perceiving unique neuro-ophthalmic visual phenomena. *Ir J Med Sci*. 2023;192(6):3139-3142. https://doi.org/10.1007/s11845-023-03315-8
34. Sanchez A, Choudhry H, Toor U, Mian S. Perception of race and sex diversity in ophthalmology by artificial intelligence. *Invest Ophthalmol Vis Sci*. 2023;64(8):5373.
35. Waikel RL, Othman AA, Patel T, et al. Generative methods for pediatric genetics education. *medRxiv*. Published online August 2, 2023. https://doi.org/10.1101/2023.08.01.23293506
36. Kenig N, Monton Echeverria J, Muntaner Vives A. Human beauty according to artificial intelligence. *Plast Reconstr Surg Glob Open*. 2023;11(7):e5153. https://doi.org/10.1097/gox.0000000000005153
37. Koljonen V. What could we make of AI in plastic surgery education. *J Plast Reconstr Aesthet Surg*. 2023;81:94-96. https://doi.org/10.1016/j.bjps.2023.04.055
38. Lim B, Seth I, Kah S, et al. Using generative artificial intelligence tools in cosmetic surgery: a study on rhinoplasty, facelifts, and blepharoplasty procedures. *J Clin Med*. 2023;12(20):6524. https://doi.org/10.3390/jcm12206524
39. Khader F, Müller-Franzes G, Tayebi Arasteh S, et al. Denoising diffusion probabilistic models for 3D medical image generation. *Sci Rep*. 2023;13(1):7303. https://doi.org/10.1038/s41598-023-34341-2
40. Adams LC, Busch F, Truhn D, Makowski MR, Aerts H, Bressem KK. What does DALL-E 2 know about radiology? *J Med Internet Res*. 2023;25:e43110. https://doi.org/10.2196/43110
41. Ali R, Tang OY, Connolly ID, et al. The face of a surgeon: an analysis of demographic representation in three leading artificial intelligence text-to-image generators. Published online. 2023. https://doi.org/10.1101/2023.05.24.23290463
42. Cho SI, Navarrete-Dechent C, Daneshjou R, et al. Generation of a melanoma and nevus data set from unstandardized clinical photographs on the internet. *JAMA Dermatol*. 2023;159(11):1223-1231. https://doi.org/10.1001/jamadermatol.2023.3521
43. DeGrave AJ, Cai ZR, Janizek JD, Daneshjou R, Lee SI. Auditing the inference processes of medical-image classifiers by leveraging generative AI and the expertise of physicians. *Nat Biomed Eng*. Published online 2023. https://doi.org/10.1038/s41551-023-01160-9
44. Zhu L, Mou W, Wu K, Zhang J, Luo P. Can DALL-E 3 reliably generate 12-lead ECGs and teaching illustrations? *Cureus*. 2024;16(1):e52748. https://doi.org/10.7759/cureus.52748
45. Javan R, Mostaghni N. AI-powered hyperrealism: next step in cinematic rendering? *Radiology*. 2024;310(1):e231971. https://doi.org/10.1148/radiol.231971
46. Ali R, Tang OY, Connolly ID, et al. Demographic representation in 3 leading artificial intelligence text-to-image generators. *JAMA Surg*. 2024;159(1):87-95. https://doi.org/10.1001/jamasurg.2023.5695
47. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;27.
48. Skandarani Y, Jodoin P-M, Lalande A. Gans for medical image synthesis: an empirical study. *J Imaging*. 2023;9(3):69.
49. Reed S, Akata Z, Lee H, Schiele B. Learning deep representations of fine-grained visual descriptions. Published online 2016.
50. Luo X, He X, Chen X, Qing L, Zhang J. DualG-GAN, a dual-channel generator based generative adversarial network for text-to-face synthesis. *Neural Netw*. 2022;155:155-167. https://doi.org/10.1016/j.neunet.2022.08.016
51. Wu F, Cheng J, Wang X, Wang L, Tao D. Image hallucination from attribute pairs. *IEEE Trans Cybern*. 2022;52(1):568-581. https://doi.org/10.1109/tcyb.2020.2979258
52. Peng J, Zhou Y, Sun X, et al. Knowledge-driven generative adversarial network for text-to-image synthesis. *IEEE Trans Multimedia*. 2022;24:4356-4366. https://doi.org/10.1109/TMM.2021.3116416
53. Wu F, Liu L, Hao F, He F, Cheng J. Text-to-image synthesis based on object-guided joint-decoding transformer. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022:18092-18101.
54. Vaswani A, Shazeer NM, Parmar N, et al. Attention is all you need. Published online 2017. https://doi.org/10.48550/ARXIV.1706.03762
55. Lampert CH, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009:951-958. https://doi.org/10.1109/CVPR.2009.5206594
56. Ronneberger O, Fischer P, Brox T. U-net: *Convolutional Networks for Biomedical Image Segmentation*. Springer; 2015:234-241.
57. Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW, Banerjee I. Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. *J Digit Imaging*. 2022;35(2):137-152. https://doi.org/10.1007/s10278-021-00556-w

58. Sorin V, Barash Y, Konen E, Klang E. Creating Artificial images for radiology applications using generative adversarial networks (GANs) - a systematic review. *Acad Radiol*. 2020;27(8): 1175-1185. https://doi.org/10.1016/j.acra.2019.12.024
59. Sun J, Jin S, Shi R, Zuo C, Jiang J. Application and prospect for generative adversarial networks in cross-modality reconstruction of medical images. *Zhong Nan Da Xue Bao Yi Xue Ban*. 2022;47(8): 1001-1008. https://doi.org/10.11817/j.issn.1672-7347.2022.220189
60. Yu B, Wang Y, Wang L, Shen D, Zhou L. Medical image synthesis via deep learning. *Adv Exp Med Biol*. 2020;1213:23-44. https://doi.org/10.1007/978-3-030-33128-3_2
61. Roster K, Kann RB, Farabi B, Gronbeck C, Brownstone N, Lipner SR. Readability and health literacy scores for ChatGPT-generated dermatology public education materials: cross-sectional analysis of sunscreen and melanoma questions. *JMIR Dermatol*. 2024;7:e50163. https://doi.org/10.2196/50163
62. Waikel RL, Othman AA, Patel T, et al. Recognition of genetic conditions after learning with images created using generative artificial intelligence. *JAMA Netw Open*. 2024;7(3):e242609. https://doi.org/10.1001/jamanetworkopen.2024.2609
63. Biggs B, Seshadri A, Zou Y, et al. Diffusion soup: model merging for text-to-image diffusion models. 2024. arXiv preprint arXiv:240608431; 2024.
64. Zhao Y, Xu Y, Xiao Z, Hou T. MobileDiffusion: Subsecond text-to-image generation on mobile devices. *arXiv preprint arXiv: 231116567*. 2023.