Viewpoint

What Does DALL-E 2 Know About Radiology?

Lisa C Adams[1,2*], MD; Felix Busch[1*], MD; Daniel Truhn[3], BSc, MD; Marcus R Makowski[4], MD, PhD; Hugo J W L Aerts[5,6], PhD; Keno K Bressem[1,5,6], MD

[1]Department of Radiology, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[2]Department of Radiology, Stanford University, Stanford, CA, United States

[3]Department of Diagnostic and Interventional Radiology, University Hospital Aachen, Aachen, Germany

[4]Department of Diagnostic and Interventional Radiology, Technical University of Munich, Munich, Germany

[5]Artificial Intelligence in Medicine Program (AIM), Mass General Brigham, Harvard Medical School, Boston, MA, United States

[6]Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, Netherlands

[*]these authors contributed equally

Corresponding Author:
Keno K Bressem, MD
Department of Radiology
Charité Universitätsmedizin Berlin
Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin
Hindenburgdamm 30
Berlin, 12203
Germany
Phone: 49 30 450 527792
Email: keno-kyrill.bressem@charite.de

Related Article:
See also: https://mededu.jmir.org/2023/1/e46885

Abstract

Generative models, such as DALL-E 2 (OpenAI), could represent promising future tools for image generation, augmentation, and manipulation for artificial intelligence research in radiology, provided that these models have sufficient medical domain knowledge. Herein, we show that DALL-E 2 has learned relevant representations of x-ray images, with promising capabilities in terms of zero-shot text-to-image generation of new images, the continuation of an image beyond its original boundaries, and the removal of elements; however, its capabilities for the generation of images with pathological abnormalities (eg, tumors, fractures, and inflammation) or computed tomography, magnetic resonance imaging, or ultrasound images are still limited. The use of generative models for augmenting and generating radiological data thus seems feasible, even if the further fine-tuning and adaptation of these models to their respective domains are required first.

(J Med Internet Res 2023;25:e43110) doi: 10.2196/43110

The Potential Impact of Generative Models in Radiology

Artificial intelligence is often seen as a potential transformer of medicine [1]. DALL-E 2 is a novel deep learning model for text-to-image generation that was first introduced by OpenAI in April 2022 [2]. The model has recently gained widespread public interest due to its ability to create photorealistic images solely from short written inputs [3-5]. Trained on billions of text-image pairs extracted from the internet, DALL-E 2 learned a wide range of representations that it can recombine to create novel images, thereby exceeding the variability it has seen in the training data, even when given implausible prompts (eg, "a painting of a pelvic X-ray by Claude Monet") [2,6-8].

DALL-E 2's powerful generative capabilities raise the question of whether these can be transferred to the medical domain to create or augment data, as medical data can be sparse and hard to come by [9,10]. Given the large number of images on which

DALL-E 2 was trained, it is reasonable to assume that the training data for DALL-E 2 included radiological images and that the model may have learned something about the composition and structure of x-ray, cross-sectional, and ultrasound images.

This viewpoint article describes how we systematically examined the radiological knowledge embedded in DALL-E 2 by creating and manipulating x-ray, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound images.

Generating Radiological Images From Short Descriptive Texts

To investigate the general radiological domain knowledge embedded in DALL-E 2, we first tested how well radiological images could be created from short descriptive texts by using the words "An X-ray of" and a 1-word description of an anatomical area. For the head, chest, shoulder, abdomen, pelvis, hand, knee, and ankle, we let DALL-E 2 synthesize 4 artificial x-ray images, which are displayed in Figure 1. For each prompt, the model produced 4 outputs, of which all are presented in order to avoid selection bias.

The coloring of the images and the general structure of the bones appeared realistic, and the overall anatomical propositions were correct, indicating the presence of fundamental concepts of x-ray anatomy. However, upon close inspection, it was noted that the trabecular structure of the bone appeared random and did not follow the course of mechanical stress as it would in real x-rays. Sometimes, smaller bones were missing, like the fibula in multiple x-rays of the knee, or multiple small bones, such as several carpal or tarsal bones, were fused into 1 bone. Nevertheless, the overall structure of the depicted anatomic area was recognizable. In rare cases, additional bones and joints were created, such as an additional index finger in a hand x-ray. The model had the greatest difficulties in generating joints correctly, as joints appeared to be fused, or the joint surfaces were distorted. Finally, the quality of the images was not what one would expect for clinic x-ray images. The collimation was largely not correct, with parts of the organ missing or the relevant regions not being centered. Examples of this are visible in Figure 1, wherein the tips of the fingers are missing in all hand x-rays, and the outer parts of the pelvis are missing in one image.

When compared to x-rays, the generation of cross-sectional images from CT or MRI was very limited (Figure 2). Although the desired anatomical structure was discernible on the images, the overall impression was predominantly nonsensical. Instead of a cross-sectional image, a photograph of an entire CT or MRI examination on x-ray film was often shown. On ultrasound images, the desired anatomic structures were not discernible; instead, all images resembled obstetric images. Nevertheless, the images showed concepts of CT, MRI, and ultrasound images, suggesting that representations of these modalities are present in DALL-E 2.

Figure 1. Examples of anatomical structures in x-ray images that were created with DALL-E 2 based on short text descriptions.
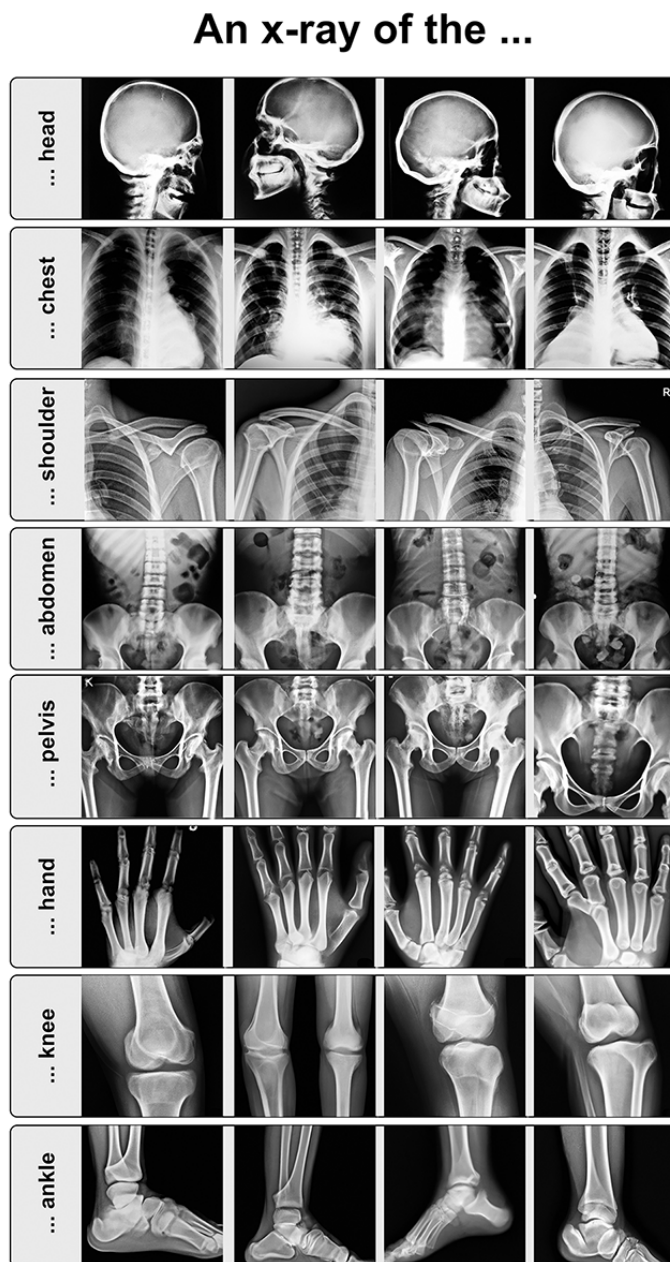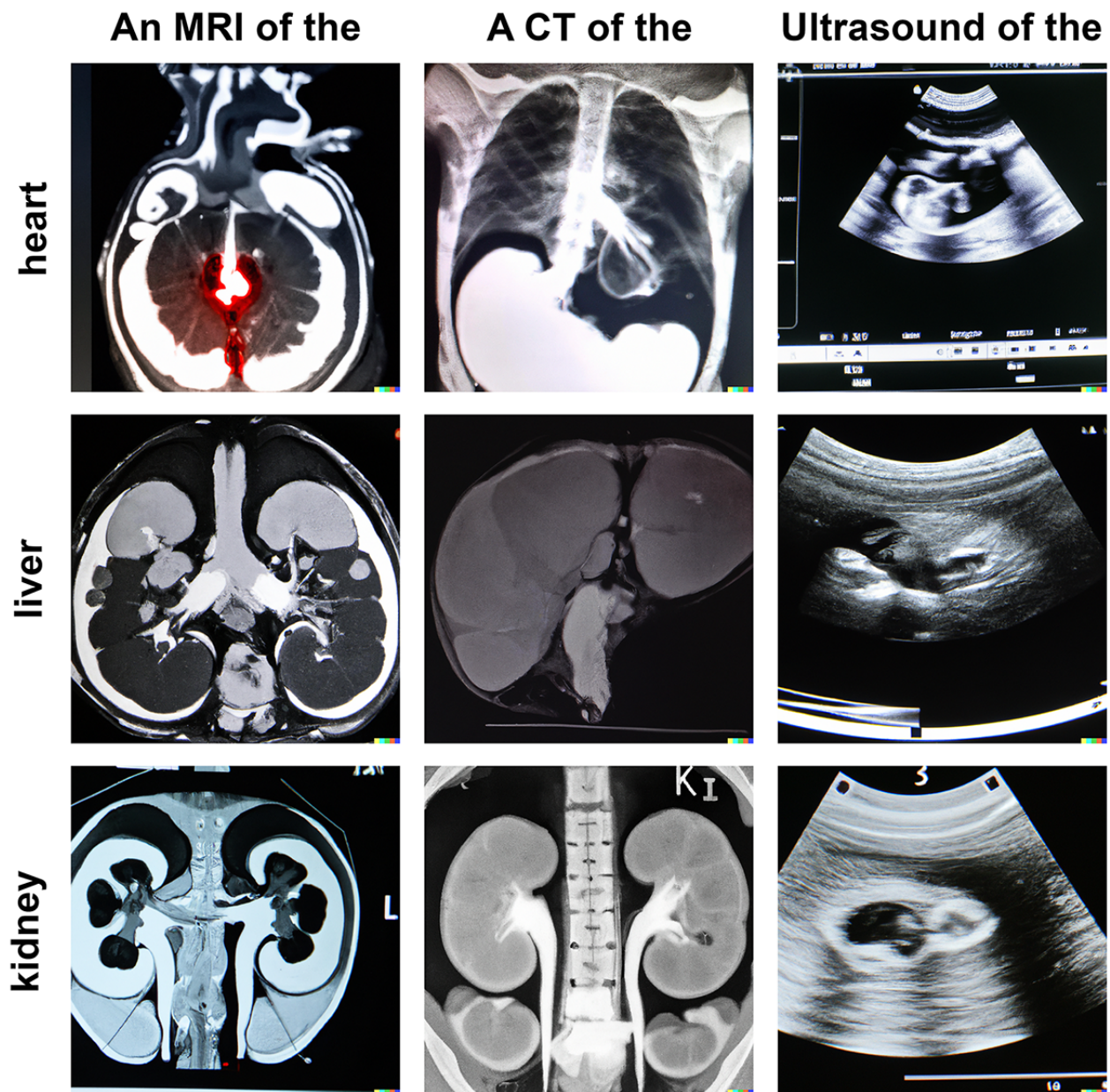
Figure 2. Examples of text-to-image–generated anatomical structures in CT, MRI, and ultrasound images created with DALL-E 2. CT: computed tomography; MRI: magnetic resonance imaging.


Changing Areas in Radiological Images While Maintaining the Anatomical Structure

To investigate the extent of DALL-E 2's radiological knowledge, we tested how well the modal can reconstruct missing areas in a radiological image (inpainting). To do this, we selected radiographs of the pelvis, ankle, chest, shoulder, knee, wrist, and thoracic spine and erased specific areas before handing the remnants to DALL-E 2 for reconstruction (Figure 3). The accompanying prompts were identical to the those used for generating the text-based images described in the *Generating Radiological Images From Short Descriptive Texts* section. DALL-E 2 provided realistic replacement images that were nearly indistinguishable from the original radiographs of the pelvis, thorax, and thoracic spine. However, the results were not as convincing when a joint was included in the image area. In the ankle and wrist images, the number of tarsal bones and the structure varied greatly from those in the original radiographs and deviated from realistic representations. For the shoulder images, DALL-E 2 failed to reconstruct the glenoid cavity and articular surface of the humerus. In one image, a foreign body was inserted into the shoulder that remotely resembled a prosthesis. Further, when reconstructing the knee image, the model omitted the patella but retained the bicondylar structure of the femur.

Figure 3. Reconstructed areas of different anatomical locations in x-rays created by using DALL-E 2. The yellow-bordered regions of the original images were erased before providing the remnant images for reconstruction.
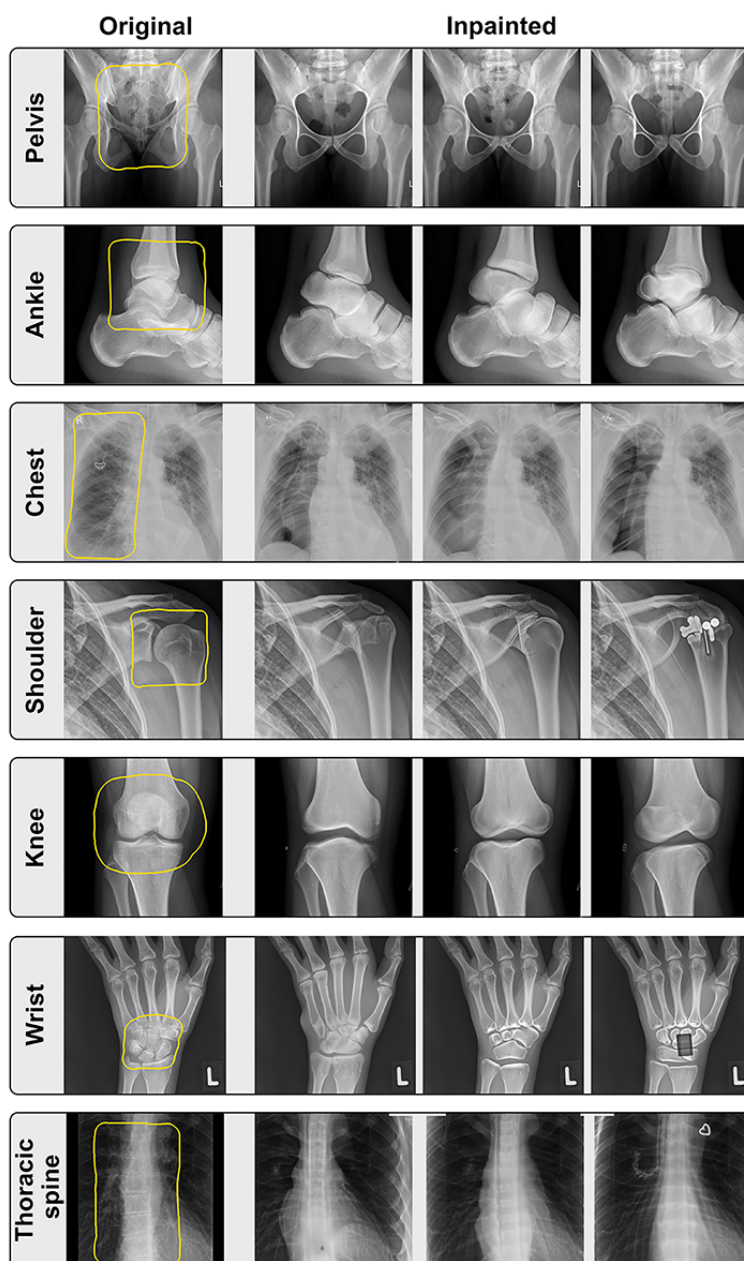

Extending Radiological Images Beyond Image Borders

Since the previous two experiments showed that DALL-E 2 could handle the basics of standard radiological images, we also wanted to investigate how well anatomical knowledge is embedded in the model. To do this, we randomly selected radiographic images of the abdomen, chest, pelvis, knee, various spinal regions, wrist bones, and hand and had DALL-E 2 extend these images beyond their boundaries, since the model needs to know both locations and anatomical proportions for this task. The accompanying prompts were selected based on the anatomical regions to be created. Again, the style of the augmented images represented a realistic representation of radiographs (Figure 4). It was possible to create a complete full-body radiograph by using only 1 image of the knee as a starting point. The greater the distance between the original image and the generated area, the less detailed the images became. Anatomical proportions, such as the length of the femur or the size of the lung, remained realistic, but finer details, such as the number of lumbar vertebrae, were inconsistent. The model generally performed best when creating anterior and posterior views, while the creation of lateral views was more challenging and produced poorer results.

Figure 4. Extending x-ray images of different anatomical regions beyond their borders by using DALL-E 2. The original x-rays are shown in yellow boxes, and the areas outside of the yellow boxes were generated by DALL-E 2.
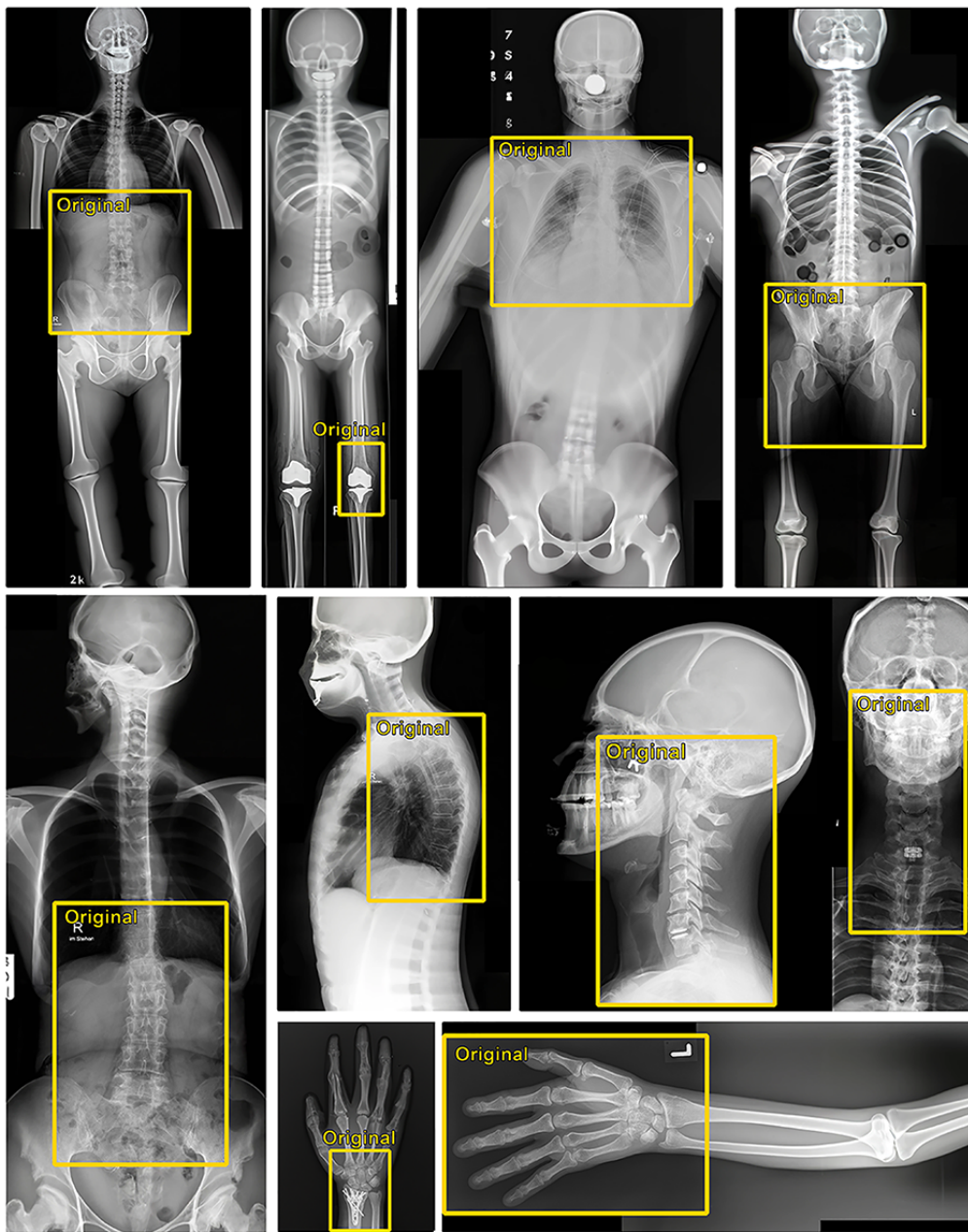
Generation of Pathological Images

The generation of pathological images to, for example, visualize fractures, intracranial hemorrhages, or tumors was limited. We tested this by generating fractures on radiographs, which mostly resulted in distorted images that were similar to those in Figure 1. In addition, DALL-E 2 has a filter that prevents the generation of harmful content. Therefore, we could not test most pathologies because words such as "bleeding" triggered this filter.

Conclusion and Outlook

We were able to show that DALL-E 2 can generate x-ray images that are similar to authentic x-ray images in terms of style and anatomical proportions. Thus, we conclude that relevant representations for x-rays were learned by DALL-E 2 during training. However, it seems that only representations for radiographs without pathologies are available or are allowed by the filter, as the generation of pathological images (eg, images of fractures and tumors) was limited. In addition, DALL-E 2's generative capabilities were poor for CT, MRI, and ultrasound images.

Access to data is critical in deep learning, and in general, the larger the data set, the better the performance of the models trained on the data set. However, especially in radiology, there is no single large database from which to create such a data set. Instead, the necessary data are divided among several institutions, and privacy concerns additionally prevent the merging of these data. Synthetic data from generative models, such as DALL-E 2, show promise for addressing these issues by enabling the creation of data sets that are much larger than

those that are currently available and greatly accelerating the development of new deep learning tools for radiology [11,12].

Previous approaches to data generation via generative adversarial networks were hampered by instabilities during training and the inability to capture sufficient scatter in the training data [13,14]. In contrast, novel diffusion-based generative models, of which DALL-E 2 is a prominent representative, solved these problems and outperformed generative adversarial networks in image generation [15].

Diffusion models are very interesting for augmenting radiological data due to their control capabilities and flexibility in data generation, as images can be modified, and pathologies can be added or removed simply by using text inputs. Nevertheless, generated images should be subjected to quality control by domain experts to reduce the risk of incorrect information from DALL-E 2 entering a generated data set. This would help to avoid possible negative effects on models trained with these data or on individuals using these data for training.

The biggest impact of these text-to-image models in radiology could be data generation for research and education. Notably, our experiments showed that DALL-E 2, in its current form, was not able to produce convincing pathological images,

probably due to the lack of such images in the training data. Radiological images containing a variety of pathologies are more likely to be provided on professional websites, may be under restricted access, or are stored as specialized data formats (eg, DICOM). Overall, these restrictions likely resulted in only a small number of images being included in the training data. In addition, the images in the data could be distorted. When looking at Figures 1 and 4, the absence of breast shadows is noticeable, leading to the assumption that all images probably show men, which makes gender bias in the model likely. Furthermore, because there is no way to ensure that the images in the training data come only from curated medical websites, incorrect information in these images may cause the model to learn incorrect concepts and thus produce biased results.

Further work is needed to explore diffusion models for medical data generation and modification [16]. Although DALL-E 2 is not publicly available, other models with a similar architecture, such as Stable Diffusion (Stability AI), were recently made available to the public [17].

Future research should focus on fine-tuning these models to medical data and incorporating medical terminology to create powerful models for data generation and augmentation in radiology research.

Data Availability

All generated research data are included in the manuscript.

Authors' Contributions

LCA and KKB conceptualized the experiments and were responsible for the methodology. LCA, FB, and KKB curated the data, conducted the formal analysis, performed the investigation, and wrote the original manuscript draft. KKB was responsible for resources and visualization. LCA, FB, DT, HJWLA, and KKB performed the validation. LCA, FB, DT, MRM, HJWLA, and KKB reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References

1. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: Scoping review. J Med Internet Res 2022 Oct 05;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
2. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents. arXiv Preprint posted online on April 13, 2022 [FREE Full text] [doi: 10.48550/arXiv.2204.06125]
3. Kather JN, Laleh NG, Foersch S, Truhn D. Medical domain knowledge in domain-agnostic generative AI. NPJ Digit Med 2022 Jul 11;5(1):90 [FREE Full text] [doi: 10.1038/s41746-022-00634-5] [Medline: 35817798]
4. Conwell C, Ullman T. Testing relational understanding in text-guided image generation. arXiv Preprint posted online on July 29, 2022 [FREE Full text] [doi: 10.48550/arXiv.2208.00005]
5. Marcus G, Davis E, Aaronson S. A very preliminary analysis of DALL-E 2. arXiv Preprint posted online on April 25, 2022 [FREE Full text] [doi: 10.48550/arXiv.2204.13807]
6. Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B, et al. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv Preprint posted online on December 20, 2021 [FREE Full text] [doi: 10.48550/arXiv.2112.10741]
7. Lyu Y, Wang X, Lin R, Wu J. Communication in human–AI co-creation: Perceptual analysis of paintings generated by text-to-image system. Appl Sci (Basel) 2022 Nov 08;12(22):11312 [FREE Full text] [doi: 10.3390/app122211312]

8.   Oppenlaender J. The creativity of text-to-image generation. 2022 Nov Presented at: Academic Mindtrek 2022: 25th International Academic Mindtrek conference; November 16-18, 2022; Tampere, Finland p. 192-202. [doi: 10.1145/3569219.3569352]

9.   Park Y, Park S, Lee M. Digital health care industry ecosystem: Network analysis. J Med Internet Res 2022 Aug 17;24(8):e37622 [FREE Full text] [doi: 10.2196/37622] [Medline: 35976690]

10. Schütte AD, Hetzel J, Gatidis S, Hepp T, Dietz B, Bauer S, et al. Overcoming barriers to data sharing with medical image generation: a comprehensive evaluation. NPJ Digit Med 2021 Sep 24;4(1):141 [FREE Full text] [doi: 10.1038/s41746-021-00507-3] [Medline: 34561528]

11. Jordon J, Szpruch L, Houssiau F, Bottarelli M, Cherubin G, Maple C, et al. Synthetic data -- what, why and how? arXiv Preprint posted online on May 6, 2022 [FREE Full text] [doi: 10.48550/arXiv.2205.03257]

12. Jordon J, Wilson A, van der Schaar M. Synthetic data: Opening the data floodgates to enable faster, more directed development of machine learning methods. arXiv Preprint posted online on December 8, 2020 [FREE Full text] [doi: 10.48550/arXiv.2012.04580]

13. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020 Nov;63(11):139-144. [doi: 10.1145/3422622]

14. Ciano G, Andreini P, Mazzierli T, Bianchini M, Scarselli F. A multi-stage GAN for multi-organ chest x-ray image generation and segmentation. Mathematics (Basel) 2021 Nov 14;9(22):2896 [FREE Full text] [doi: 10.3390/math9222896]

15. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. 2021 Presented at: 35th Conference on Neural Information Processing Systems (NeurIPS 2021); December 6-14, 2021; Virtual-only conference.

16. Tang Y, Tang Y, Zhu Y, Xiao J, Summers RM. A disentangled generative model for disease decomposition in chest x-rays via normal image synthesis. Med Image Anal 2021 Jan;67:101839. [doi: 10.1016/j.media.2020.101839] [Medline: 33080508]

17. Buehler MJ. Generating 3D architectured nature-inspired materials and granular media using diffusion models based on language cues. Oxf Open Mater Sci 2022 Nov 11;2(1):itac010 [FREE Full text] [doi: 10.1093/oxfmat/itac010] [Medline: 36756638]

Abbreviations

CT: computed tomography
MRI: magnetic resonance imaging