



# Updated Primer on Generative Artificial Intelligence and Large Language Models in Medical Imaging for Medical Professionals

Kiduk Kim<sup>1</sup>, Kyungjin Cho<sup>2</sup>, Ryoungwoo Jang<sup>3</sup>, Sunggu Kyung<sup>2</sup>, Soyoung Lee<sup>2</sup>, Sungwon Ham<sup>4</sup>, Edward Choi<sup>5</sup>, Gil-Sun Hong<sup>6\*</sup>, Namkug Kim<sup>1,6\*</sup>

<sup>1</sup>Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

<sup>2</sup>Department of Biomedical Engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

<sup>3</sup>Coreline Soft Co., Ltd., Seoul, Republic of Korea

<sup>4</sup>Healthcare Readiness Institute for Unified Korea, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Republic of Korea

<sup>5</sup>Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

<sup>6</sup>Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

The emergence of Chat Generative Pre-trained Transformer (ChatGPT), a chatbot developed by OpenAI, has garnered interest in the application of generative artificial intelligence (AI) models in the medical field. This review summarizes different generative AI models and their potential applications in the field of medicine and explores the evolving landscape of Generative Adversarial Networks and diffusion models since the introduction of generative AI models. These models have made valuable contributions to the field of radiology. Furthermore, this review also explores the significance of synthetic data in addressing privacy concerns and augmenting data diversity and quality within the medical domain, in addition to emphasizing the role of inversion in the investigation of generative models and outlining an approach to replicate this process. We provide an overview of Large Language Models, such as GPTs and bidirectional encoder representations (BERTs), that focus on prominent representatives and discuss recent initiatives involving language-vision models in radiology, including innovative large language and vision assistant for biomedicine (LLaVa-Med), to illustrate their practical application. This comprehensive review offers insights into the wide-ranging applications of generative AI models in clinical research and emphasizes their transformative potential.

**Keywords:** Artificial intelligence; Generative artificial intelligence; Large language model; Synthetic data; Medical imaging

## INTRODUCTION

The advent of Chat Generative Pre-trained Transformer (ChatGPT), a powerful tool developed by OpenAI, has garnered interest in various generative artificial intelligence (AI) models in the field of medicine. Generative AI, which

includes Generative Adversarial Networks (GANs) [1,2], diffusion models [3-5], and Large Language Models (LLMs) [6-10], has shown tremendous potential for a wide range of medical applications.

Medical imaging has emerged as a prominent area that can be explored using generative models. GANs and

**Received:** August 28, 2023 **Revised:** November 27, 2023 **Accepted:** December 28, 2023

\*These authors contributed equally to this work.

**Corresponding author:** Gil-Sun Hong, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Republic of Korea

• E-mail: hgs2013@gmail.com

**Corresponding author:** Namkug Kim, PhD, Department of Convergence Medicine, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Republic of Korea

• E-mail: namkugkim@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

diffusion models have been used in studies focusing on image reconstruction and quality enhancement [11]. The paramount importance of maintaining privacy in medical research has also led to the use of synthetic data. Owing to their ability to produce synthetic medical data that mirror real-world characteristics, generative models offer innovative solutions for data privacy, thereby enabling researchers to conduct studies without compromising patient confidentiality.

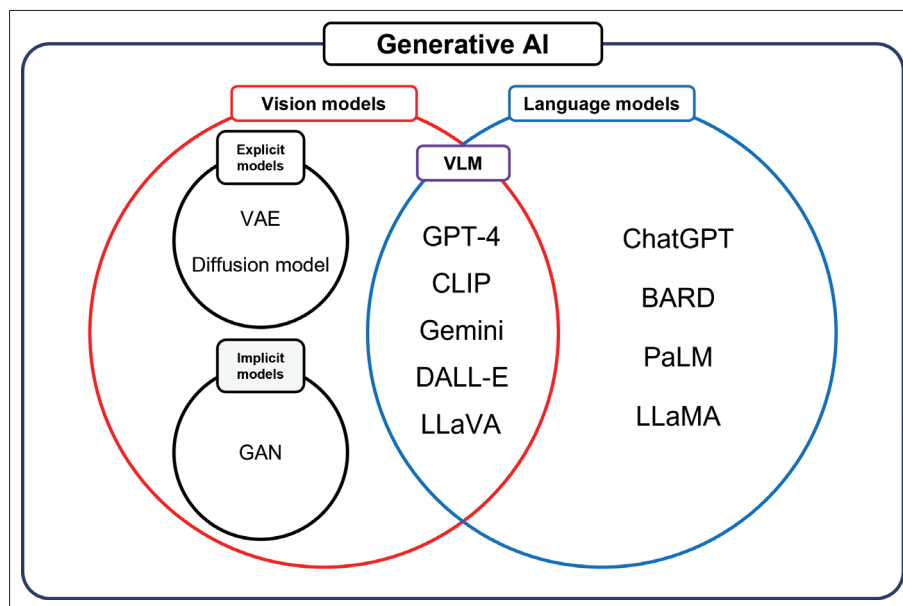
The vision-language model (VLM) and LLMs, such as ChatGPT, have accelerated the developmental pace and facilitated the development of numerous applications in the medical field that were previously unimaginable. The increase in the use of AI-driven models in the medical field has necessitated a comprehensive understanding of generative AI models among researchers and practitioners, including general radiologists seeking to leverage these advancements.

Therefore, this review aims to provide a concise overview of the fundamental principles and applications of generative AI models, with a particular focus on medical imaging. This review explores the introduction of basic models, such as variational autoencoders (VAEs), GANs, diffusion models, and their variants, and the underlying mechanisms that drive their generative capabilities. Furthermore, this review also showcases how these models combine their capabilities with the power of LLMs to understand images.

## Overview of Common Generative AI Models: VAE, GAN, Diffusion Models, LLMs, and VLMs

In contrast to deterministic models, such as regression models and classifiers, wherein the outcome is completely determined by the parameter and initial input values, generative models incorporate randomness or unpredictability while predicting the answers. Consequently, generative models are referred to as probabilistic or stochastic models. Probabilistic models incorporate probability (chance) to address uncertainty and consider the likelihood of obtaining different outcomes during the prediction process. Stochastic models inherently involve randomness or unpredictability in their prediction processes. These generative models can be categorized according to the type of data process (i.e., vision, language, or both) or the manner in which they train and generate samples (i.e., explicit and implicit models). Explicit models are probabilistic generative models that define the probability distribution of the modeled data. In contrast, implicit models generate data directly via stochastic or random processes without explicitly defining the probability distribution. Figure 1 illustrates the categories of commonly used generative AI models.

Vision generative models, such as VAEs [12], GANs [1,2,13], and diffusion models [3-5], are defined as models that only process images. Figure 2 depicts the typical architecture of



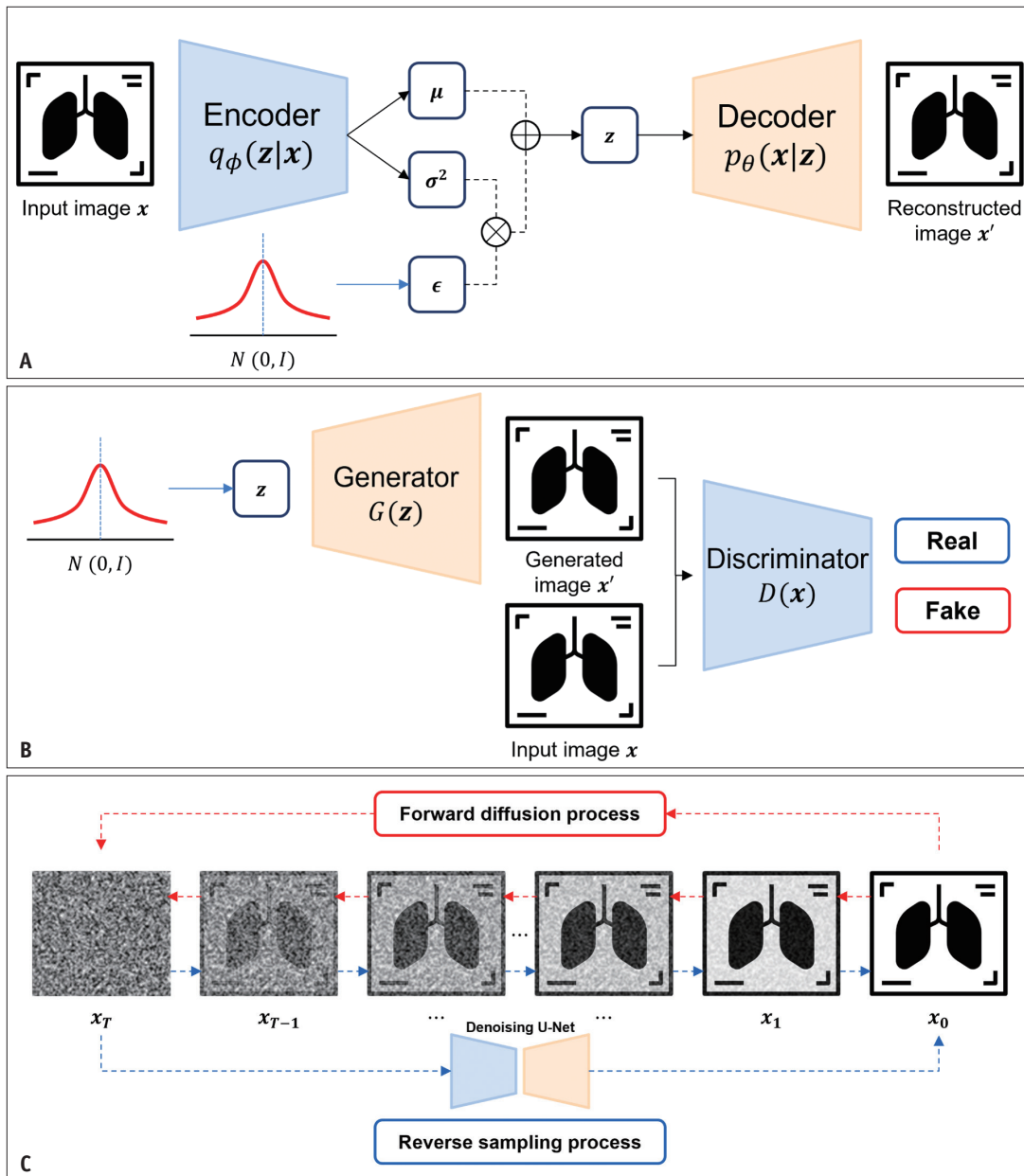
**Fig. 1.** Categories and examples of generative AIs. AI = artificial intelligence, VLM = vision-language model, VAE = variational auto-encoder, GPT = generative pre-trained transformer, CLIP = contrastive language-image pretraining, PaLM = pathways language model, LLaVA = large language and vision assistant, LLaMA = large language model Meta AI, GAN = generative adversarial network

vision-generative models. Remarkable progress in natural language processing (NLP) has facilitated the processing and generation of a massive amount of data by language-generative models, such as ChatGPT, BARD, pathways language model (PaLM) [14], and large language model Meta AI (LLaMA) [15]. Recent advances in multimodal processing in deep learning have also facilitated the processing of natural language and computer vision data by large models. Consequently, AI models can generate language from images, images from language, or both. Examples of VLMs include

GPT-4 [16], DALL-E [17,18], and large language and vision assistant (LLaVA) [19].

### Variational Auto-Encoder (VAE)

VAEs are generative models that employ an encoder-decoder architecture with a prior distribution (existing distribution). The encoder maps each input image onto a latent space. The encoded latent feature is subsequently used by the decoder to generate an image. VAEs are



**Fig. 2.** Typical architectures of generative models for medical images. **A:** Variational auto-encoder. **B:** Generative adversarial network. **C:** Diffusion model.  $x$  = input image,  $z$  = latent space,  $N$  = normal distribution,  $\epsilon$  = noise,  $x'$  = reconstructed image,  $T$  = timestep

trained by approximating the distribution of the encoded latent features to a known distribution (e.g., a normal distribution) and reconstructing the image to resemble the given input. An auto-encoder [20] possesses a similar architecture; however, it focuses on the efficient learning of latent representations rather than image generation.

### Variants of VAE Models

VAEs, which are a class of generative models based on the principles of variational inference, have several variants. Conditional VAEs (CVAEs) [21] are an extension of VAEs wherein the encoder and decoder are conditioned on additional information, such as labels or other data. VAEs are particularly useful for tasks such as conditional image generation (e.g., generating images of a particular class). VAEs with arbitrary conditioning (VAEACs) [22] enable flexible conditioning (i.e., conditioning on arbitrary subsets of observed variables). Consequently, VAEACs have been used for various tasks such as inpainting, denoising, or feature prediction. Disentangled VAE (beta-total correlation VAE,  $\beta$ -TCVAE) [23] introduces an additional term in the loss function that facilitates the learning of disentangled representations by minimizing the total correlation between different latent variables. Hierarchical VAEs [24] are characterized by a hierarchy of latent variables, with each level capturing a different level of abstraction in the data. Hierarchical VAEs can model complex data distributions more effectively and capture abstract representations at higher levels of hierarchy. Sequential VAEs, such as deep recurrent attentive writers (DRAW) [25] and variational recurrent neural networks (VRNN) [26], can handle sequential data, such as text or time series.

## Generative Adversarial Network (GAN)

GANs comprise two competing networks: a generator, which forges a realistic fake image from a given latent feature, and a discriminator, which distinguishes fake images from real images. The generator attempts to deceive the discriminator via adversarial training, thereby improving image generation. Mode collapse, a phenomenon characterized by the production of a limited variety of samples by the generator, and training instability are the primary limitations of standard GANs.

### Variants of GAN Models

Different variants of GANs, such as deep convolutional

GAN (DCGAN), conditional GAN (cGAN), progressive growing GAN (PGGAN), style based GANs (StyleGAN), cycle-consistent GAN (CycleGAN), and StarGAN, are characterized by unique characteristics and applications.

DCGANs integrate convolutional neural networks (CNNs) into the architecture of GANs [27]. Transposed convolutional layers are used by the generator to produce images, and common convolutional layers are used by the discriminator to distinguish between the real and generated images. This marks a significant step forward, as most GANs developed prior to DCGANs were based on fully connected layers. Compared with standard GANs, DCGANs yield a significant improvement in the stability of training GANs and facilitate the generation of higher-quality images.

In contrast to standard GANs that generate data from random noise, cGANs [28] generate data conditioned on additional information, such as class labels, data from other modalities, or text descriptions. This conditionality enables the generation of targeted types of images, such as images of a specific class.

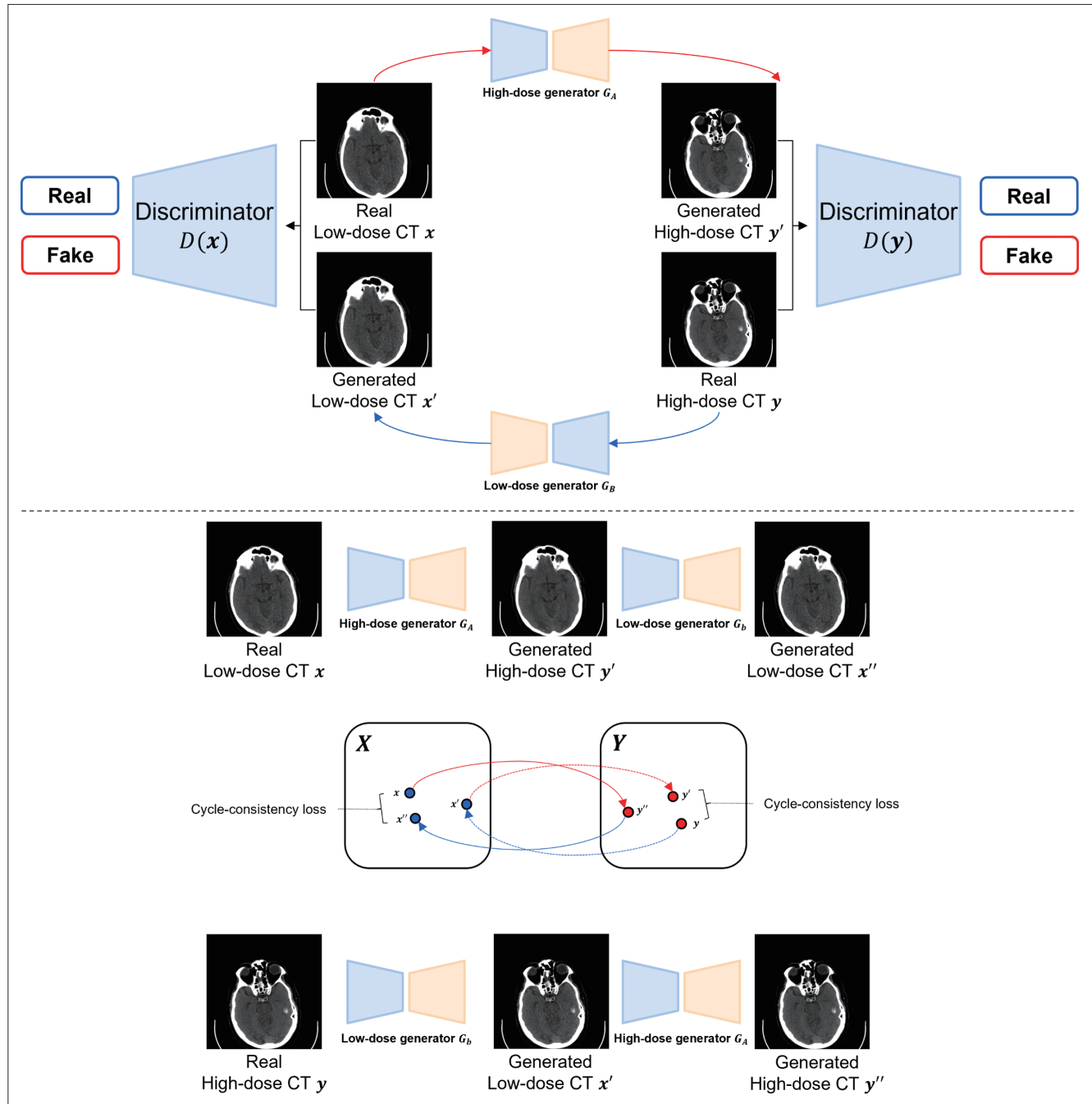
PGGANs [29] progressively train generator and discriminator networks, starting from low-resolution images and gradually increasing to higher resolutions. This approach enables the networks to learn large-scale structures initially and then learn finer details progressively as the training progresses. PGGANs achieve more stable training compared with that of a high-resolution GAN from the beginning by focusing on lower resolutions initially and then gradually increasing the complexity.

StyleGANs [2,30,31] introduce a novel architecture, such as a style-based generator, that facilitates unprecedented control over the style and content of the generated images. This key innovation in StyleGANs facilitates separate control over high-level attributes (such as the contour and shape of the brain) and stochastic variations (such as brain sulci) in the generated images. StyleGANs [2] can also introduce a mapping network that transforms a latent code (random input) into an intermediate latent space. This intermediate space can capture the 'style' of the generated image. In addition to style, noise inputs can be added at various layers to introduce stochastic variations that are not controlled by the latent code. Improved versions, such as StyleGAN2 [30] and StyleGAN3 [31], have been developed in addition to the original StyleGAN.

CycleGAN [32], which was specifically designed for image-to-image translation where paired examples are not available, uses two-generator networks. Generator  $G_x$  learns to map

from domain X to domain Y, whereas generator  $G_Y$  learns reverse mapping from domain Y to domain X. CycleGAN also comprises two discriminators,  $D_X$  and  $D_Y$  (Fig. 3). Cycle consistency loss is a crucial component of CycleGAN that ensures the translation of an image from domain X to domain Y and then back to domain X such that the generated image

resembles the original image from domain X (and vice versa for domain Y). This component acts as a proxy for the paired training data. Moreover, its ability to work without paired examples facilitates its incorporation into a wide range of real-world scenarios where paired training data are scarce or difficult to obtain. For instance, CycleGANs can convert the



**Fig. 3.** CycleGAN architecture to generate high-dose CT images from low-dose CT using an unpaired dataset. CycleGAN = cycle-consistent generative adversarial network, CT = computed tomography, D = discriminator,  $G_A$  = high-dose CT generator from low-dose CT,  $G_B$  = low-dose CT generator from high-dose CT,  $x$  = low-dose CT image,  $y$  = high-dose CT image,  $X$  = domain of low-dose CT,  $Y$  = domain of high-dose CT, ' = generated image from real image, '' = generated image from generated image

reconstruction kernel of computed tomography (CT) images or denoise low-dose CT [33-38].

StarGAN [39], a versatile GAN that can perform multiple image-to-image translations simultaneously, can handle multiple domains when translating between multiple medical imaging modalities, such as multi-contrast magnetic resonance (MR) images and multi-view echocardiography [40,41]. StarGAN is particularly useful in scenarios encompassing diverse datasets and imaging equipment.

## Diffusion Model

The diffusion model, a newer generative model, uses noise distribution and operates in forward and reverse processes based on a Markov chain, a discrete time-stochastic process with Markov property. This property dictates that the future state of the process is determined solely by its current state and is independent of its history. The transition from one state to another is determined by the probability associated with the current state. Owing to this property, a Markov chain can efficiently describe the transition between two states over time (or steps) by multiplying the initial probability by the probabilities of subsequent steps. Thus, it is a concise and easy method that can be used to determine dynamic changes in the data. Random noise (i.e., Gaussian noise) is added to a given image in multiple scheduled steps in the forward process of a diffusion process. In contrast, the model denoises the given image and predicts the image from the previous step in the reverse process. The model is trained by predicting “the noise” from the given image, and images are generated by denoising a random noise step by step (a Markov chain). Figure 4 presents an example of the generation of three-dimensional medical images from two-dimensional slices using the mask inpainting technique of a diffusion model sampler.

### Variants of Diffusion Models

Variants of diffusion models include score-based generative models, denoising diffusion probabilistic models (DDPMs) [3], and denoising diffusion implicit models (DDIMs). The “score,” which refers to the gradient of the log probability density of the data in score-based models [5], serves as a guiding factor for the forward (where noise is added to the data) and reverse (where the data are recovered from noise) processes utilizing stochastic differential equations.

DDPMs are based on the principle of a diffusion process, which is a Markov chain that gradually adds noise to data

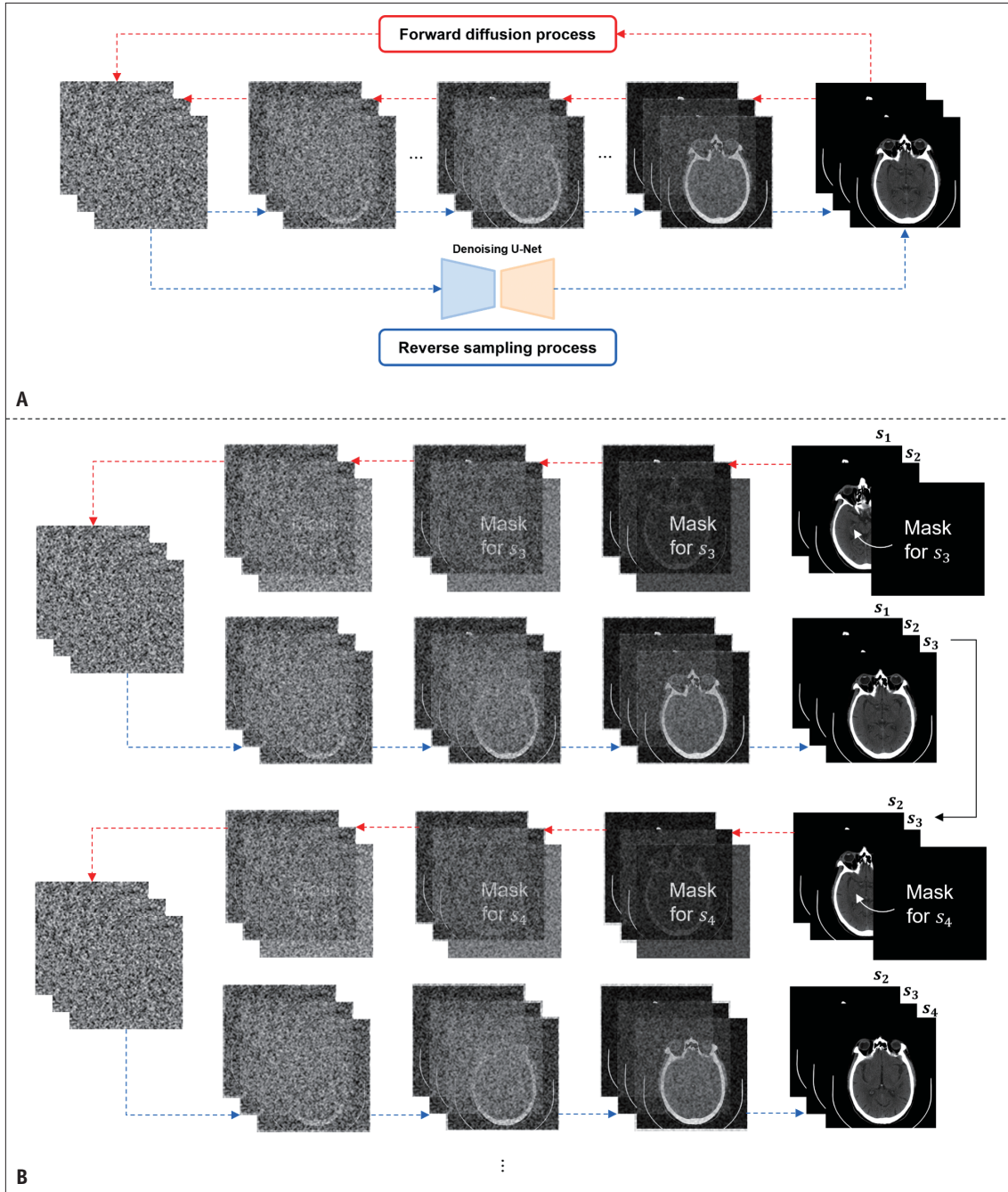
over a series of steps. This process transforms data into a simple known distribution (typically Gaussian noise). Forward diffusion is a fixed process that is not learned, whereas the reverse process is learned by a neural network trained to predict the noise added at each step of the forward process. The model can reconstruct the original data from noise via iterative denoising.

DDIMs modify the traditional diffusion process to facilitate efficient and deterministic generation of samples [4]. In contrast to DDPMs, which explicitly model noise at each step of the diffusion process, DDIMs use an implicit modeling approach. Consequently, DDIMs can define a deterministic trajectory for the reverse process (from noise to data) without explicitly modeling the noise distribution at each step. The deterministic non-Markovian reverse process of DDIMs is one of its unique features. The reverse process involves the addition of a small amount of random noise at each step in DDPMs, making the process inherently stochastic. In contrast, DDIMs follow a deterministic path. Consequently, the output remains the same, given a starting noise. DDIMs can generate samples more rapidly, while ensuring minimal changes in the quality of the generated image. This is particularly beneficial for reducing computational overhead.

## Large Language Model (LLM)

LLMs, such as ChatGPT and GPT-4 [16], are based on transformer architecture [42]. The transformer, the core component of LLMs, enables these models to understand and generate human-like text. Words are processed simultaneously in the transformer, rather than one after the other, which makes it a great tool for understanding the context of a language. Transformers use a mechanism known as “attention” to weigh the relevance of different words when generating responses or predictions. LLMs train on vast amounts of text data and scale the model architecture to a larger size.

GPT is an LLM that uses a transformer decoder, a specific part of the transformer architecture. As it comprises a decoder-only architecture, it resembles an expert chef who does not have to learn a specific recipe (encoder) since they have already learned hundreds of recipes. Consequently, GPT-based LLMs [6-8,16], such as ChatGPT and GPT-4, can effectively generate human-like text by combining the power of the transformer decoder with the broad knowledge learned from extensive training data. GPT can generate texts based on a description or a single or few examples. Figure 5 presents

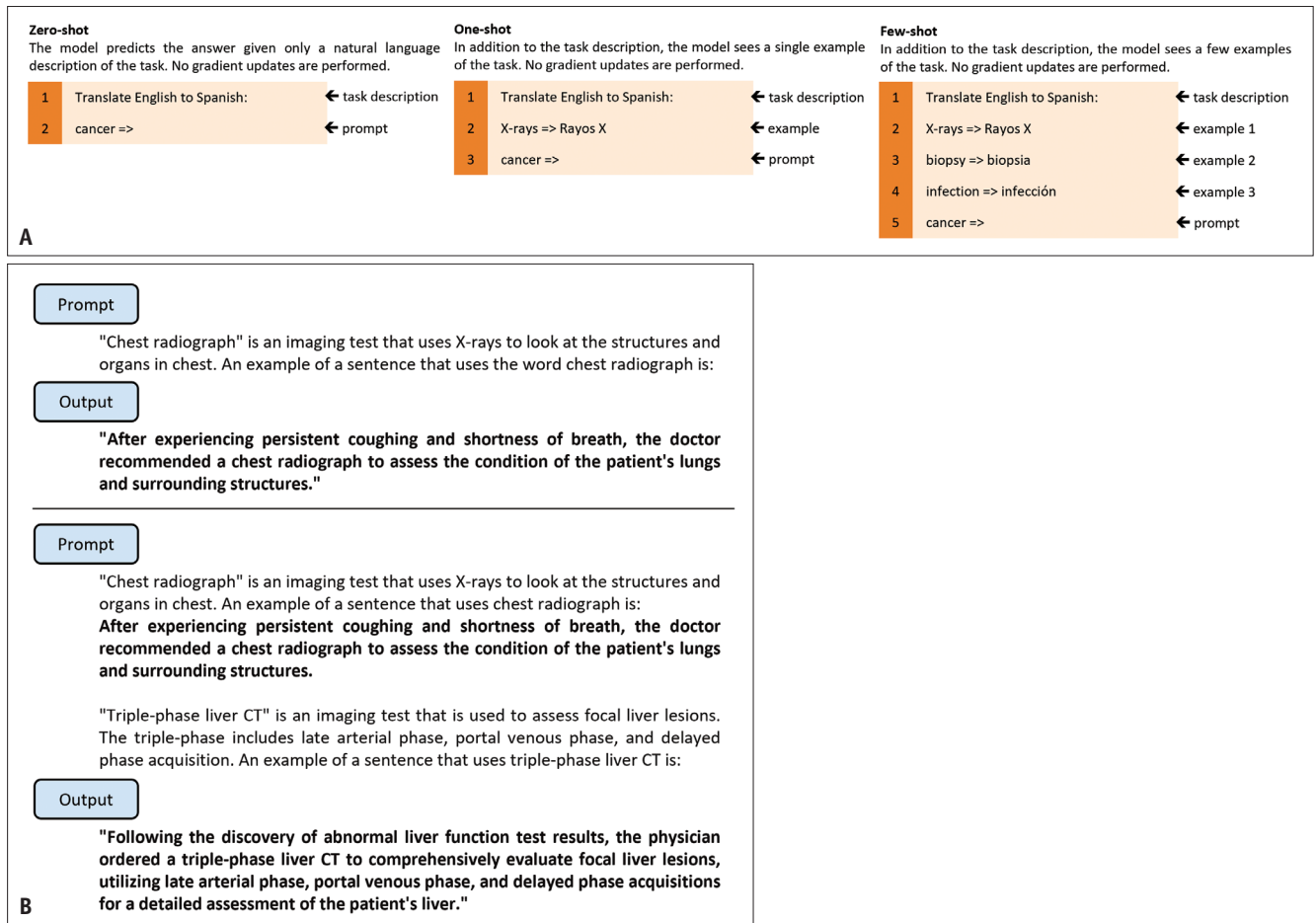


**Fig. 4.** 2D slice-wise generation of 3D CT volume using the mask inpainting technique of the denoising diffusion probabilistic model sampler. **A:** Training 3D diffusion model. **B:** 3D slice generation using diffusion sampling. D = dimensional, CT = computed tomography,  $s_i$  =  $i^{\text{th}}$  slice of CT image, Mask for  $s_i$  = empty mask for generative inpainting to generate  $s_i$  slice

zero-shot, one-shot, and few-shot generation examples of GPT.

Bidirectional Encoder Representations from Transformers (BERTs) [10] are pre-trained models that use the bidirectional encoder of the transformer architecture, which facilitates the consideration of the left and right contexts during training. BERT is trained on a process known as the masked language model, which masks random words in a sentence. The model

predicts the masked words based on the context provided by the surrounding sentences and words. This process has enabled BERT to capture deep contextual representations, resulting in superior performance in various NLP tasks. BERT is a powerful tool for NLP tasks, such as language understanding, sentiment analysis, and question-answering systems, owing to its ability to understand context.



**Fig. 5.** Description of zero-shot, one-shot, and few-shot generation and prompting. **A:** Description of zero-shot, one-shot, and few-shot generation. **B:** Prompting examples.

GPT and BERT are built on the transformer architecture and utilize attention mechanisms, which significantly improve NLP tasks. These models have facilitated the completion of more accurate and context-aware language-modeling tasks, making them valuable assets in the domain of NLP. The capabilities of these models have given rise to new avenues for various applications, from the generation of coherent text to understanding and processing languages in a meaningful and contextually relevant manner. LLMs can help create a patient-friendly language for reports and discussions, improve communication with patients, and enhance the understanding of medical conditions and imaging results among patients [43-45].

### Large Language Model to Vision-Language Model

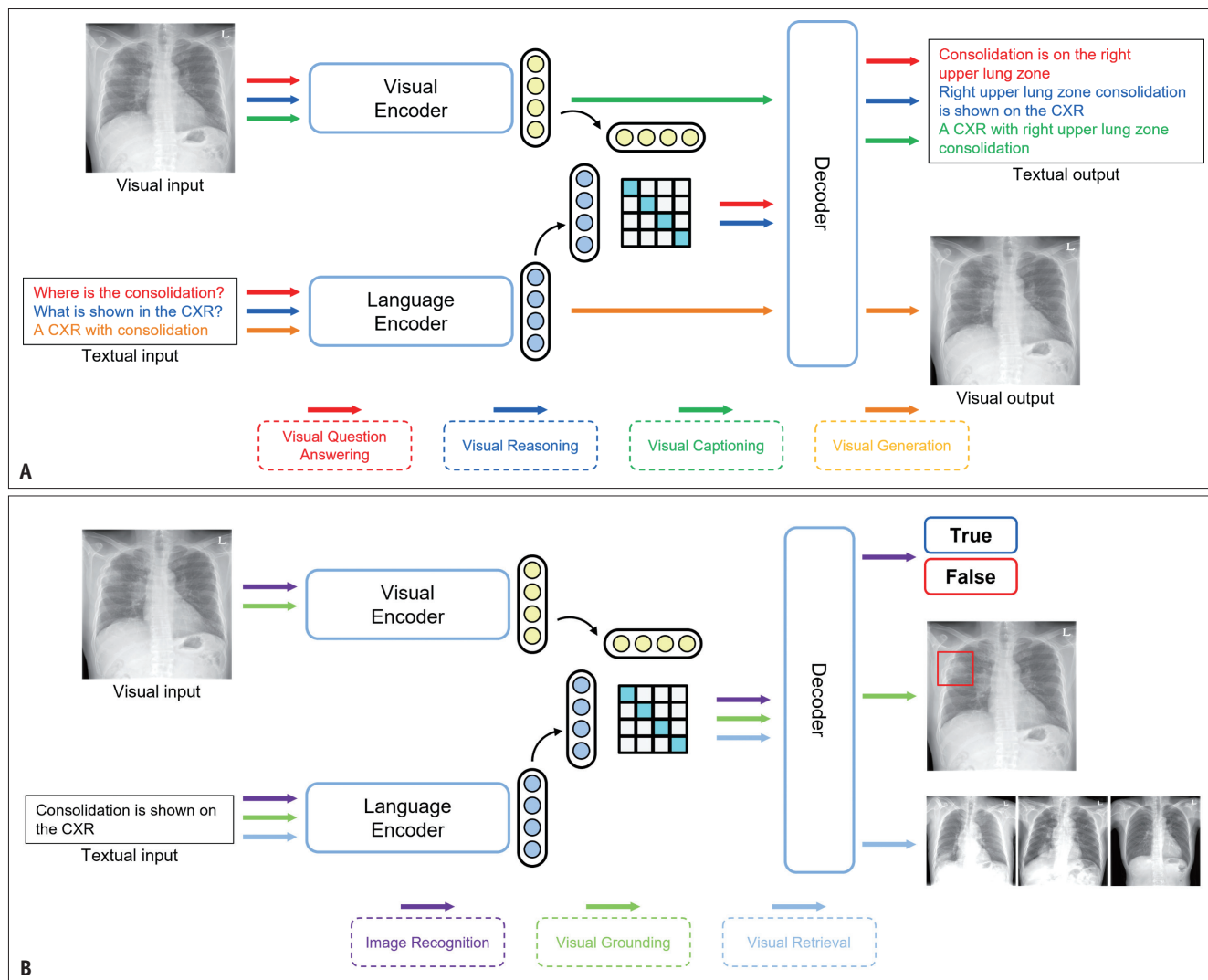
VLM [46] has been introduced as an AI model that combines the power of natural language understanding with images or visual understanding in recent years. VLMs function by learning from a large amount of text and image

data in a manner similar to that of a student learning by reading books and text accompanied by pictures. Owing to their ability to quickly analyze medical reports and images simultaneously, these models can be used as aids by radiologists.

The tasks performed in the field of VLM, which is a multidisciplinary field that combines computer vision and NLP, can be classified into two main categories: generation and perception.

Generation tasks can be grouped into four categories: visual question answering (VQA), visual reasoning, visual captioning, and visual generation. Figure 6A presents the generation tasks in VLM. AI models are presented with a visual input (an image or video) and a question related to the input in VQA [47]. The AI model provides correct responses based on its understanding of the questions and visual input. Visual reasoning requires the deduction of cognitive insights or commonsense knowledge from images by the AI models [48]. AI models create relevant and





**Fig. 6.** Generation and perception tasks of language-vision models. **A:** Generation language-vision models. **B:** Perception language-vision models. CXR = chest radiograph

descriptive captions for the visual inputs provided in visual captioning [49]. Visual generation involves the generation of visual output from a given textual input [50].

Perception tasks can be grouped into three categories: image recognition, visual grounding, and visual retrieval. Figure 6B presents the perception tasks in VLM. Contrastive language-image pretraining (CLIP) [51] and a large-scale image and noisy-text embedding (ALIGN) [52] have transformed traditional image recognition tasks into language-vision tasks via image recognition, thereby enabling the recognition of unseen concepts by AI. Visual grounding involves the prediction of the bounding box that corresponds to a text query within an image [53,54]. Image-text retrieval involves the retrieval of the most relevant text or image from a large corpus based on the query provided [55].

General segmentation with prompting uses textual prompts to guide the AI in the identification and outlining of specific parts of an image. The Segment Anything Model (SAM) developed by Meta [56] has exhibited notable performance in this area. This approach can also be applied to the domain of medical imaging [57-61], making it a promising avenue for future research.

### Outlook of Foundation Models in Radiology

A foundation model is a large model that is pre-trained on a vast amount of data. Foundation models can be used for diverse downstream tasks via fine-tuning or zero-shot methods [62]. Representative language foundation models include BERT [10] and the GPT series [6-8]. Several multi-modal foundation models of VLM, such as DALL-E [17,18],

Flamingo [63], and Florence [64], have also been introduced.

A research team from Microsoft recently reported a significant breakthrough in the biomedical domain with the introduction of LLaVA for BioMedicine (LLaVA-Med), a VLM [65]. A research team from Google introduced its counterpart, Med-PaLM, around the same time [66]. These foundation medical VLMs leverage the ability to understand images and interpret text to perform various tasks in the domain of biomedical imaging. The primary objective of these models is to assist in answering open-ended research questions related to biomedical images. To this end, the models learned from a vast biomedical figure-caption dataset extracted from PubMed Central, a large biomedical literature database. The training of LLaVA-Med was performed using a curriculum learning method, wherein learning tasks are presented in a sequence of increasing difficulty, in a manner similar to how humans learn, starting with simple concepts and progressing to more complex ones. The model initially focused on aligning biomedical vocabulary using figure-caption pairs. Subsequently, the model learned to understand open-ended conversational semantics using the data generated by GPT-4, another powerful AI language model. This training process enabled the model to acquire biomedical knowledge gradually, in a manner similar to how a layperson learns. This led to the development of an AI system capable of answering inquiries regarding biomedical images and open-ended research questions. These foundation models of VLMs perform well on standard biomedical VQA datasets [65,66]. The advantage of this system lies in its multimodal approach, as it can comprehend textual and visual information.

Attempts have been made to construct such foundation models in the field of radiology. One group of researchers trained a vision foundation model on 100 million medical images, including radiographs, CT images, MR images, and ultrasound images [67]. Another group of researchers trained a self-supervised network on 4.8 million chest radiographs [68]. However, the networks were not scalable despite these studies training their networks on a vast amount of data and demonstrating their diverse utility. A foundation VLM was implemented in the field of radiology in a recent study [69].

## Potential and Application of Generative Models in Clinical Imaging Research

Generative models have demonstrated versatility in numerous tasks, such as denoising, image reconstruction,

and vital image analysis. Moreover, they have also been utilized in a technique called “inversion” to transform real data, such as images, into a latent space representation. These models have emerged as a promising approach that can address multiple challenges in medical research by generating medical data.

### Denoising, Image Reconstruction, Inter-Modality Synthesis, and Imaging Analysis Tasks

Promising applications of GAN in the field of radiology include image quality improvement via post-processing and faster image acquisition. Radiation exposure and the increased risk of developing cancer associated with CT image acquisition are obstacles faced in the field of radiology. Deep learning-based image reconstruction, particularly using GANs, may be a potential solution to overcome this obstacle. Successful applications of GANs include CT denoising [70,71], artifact reduction [72], and improving the accuracy of radiation therapy planning [73,74]. MRI modalities often require long acquisition times. Deep learning techniques, including GANs, have been used to enable high-quality image reconstruction from undersampled k-space data [75-77], effectively speeding up MRI acquisition and improving image quality. In addition, GANs have also been utilized to convert low-magnetic-field MR images into high-magnetic-field MR images [78].

Intermodality synthesis, which has been used extensively to optimize imaging processes, involves converting images between different modalities or sequences [79]. GANs play a crucial role in this process and have been used to synthesize CT images from MR images [80,81], thereby reducing radiation exposure and diminishing the requirement for additional image acquisitions. GANs can also compensate for missing MRI sequence data [82] and support training across multiple modalities [83], thereby facilitating image analysis across various modalities and sequences.

GANs are versatile and effective tools for advancing medical imaging and analysis. Moreover, they have effectively improved the deep learning performance for various radiology tasks, including lesion detection, organ segmentation, and the prediction of patient outcomes, via data augmentation [84-89]. GANs have also been used in image registration to yield more accurate results. They have been used successfully in MR-to-transrectal ultrasound image registration and image registration of MR and CT images for thoracic and abdominal organs [90,91]. Furthermore, GANs have also been used to identify abnormal lesions in medical

images by learning the data distribution of normal images via unsupervised or semi-supervised learning [92-97]. The progression of Alzheimer's disease has been modeled using GANs via the utilization of MRI data to predict disease changes over longitudinal examinations [98].

Diffusion models have been used effectively in various CT imaging applications, such as CT denoising [99-102], CT kernel conversion [103], and accelerating MRI acquisition time [104-106]. Diffusion models have also demonstrated potential for intermodality synthesis, notably generating synthetic CT images from MR images [107-109]. Moreover, these models have been applied to other domains of radiology, such as automatic lesion and organ segmentation [110-113], image registration [114], and anomaly detection [115-119].

### Inversion of Generative Models

Certain real data (e.g., images) can be inverted into a latent space using a technique known as inverse mapping of generative models or inversion of generative models [120]. The inverted latent features are manipulated and edited subsequently [121-123]. Figure 7 presents an example of the manipulation of an inverted latent feature and the editing of the images using the generative model inversion technique. GANs have been used for inversion in the medical domain. Ren et al. [124] used GAN inversion to generate tumor-like stimuli with specific shapes, sizes, and realistic textures in mammograms in a controlled manner. Fetty et al. [125] reported that GAN inversion can be used to manipulate the imaging modality (MRI to CT and vice versa) and sex of the patient. Lee et al. [126] demonstrated that GAN inversion can be used to control the presence and severity of a disease by manipulating the latent space. These findings indicate that GAN inversion can address important issues of deep learning in medical imaging and help develop other deep-learning networks by providing images of diverse disease severities.

### Solving Clinical Research Challenges using Generative Model-Based Synthetic Data

Stringent regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe, have restricted the sharing of sensitive patient data, thereby impeding the creation of comprehensive datasets for research [127,128]. Synthetic data sharing has safeguarded patient privacy and facilitated research collaboration [129].

Imbalances in medical datasets, particularly in datasets of underrepresented rare diseases, can skew machine-learning predictions. Synthetic data has been used to rectify this imbalance using methods such as the synthetic minority oversampling technique (SMOTE) by addressing class imbalances [130]. Health inequities are amplified by real-world data bias caused by factors such as age, sex, and socioeconomic status. DEbiasing CAusal Fairness (DECAF) mitigates this bias by leveraging causal structures to synthesize fairer data [131]. Synthetic data, which have been used in clinical simulation tests for AI models in diverse settings, aid in identifying and rectifying potential errors before deployment, thereby minimizing "alert fatigue" and bolstering trust in AI models [132]. Synthetic control arms utilizing external patient-level data are efficient and ethical alternatives that can be used in clinical trials to save resources [133,134].

### Pitfalls and Limitations of Generative AI

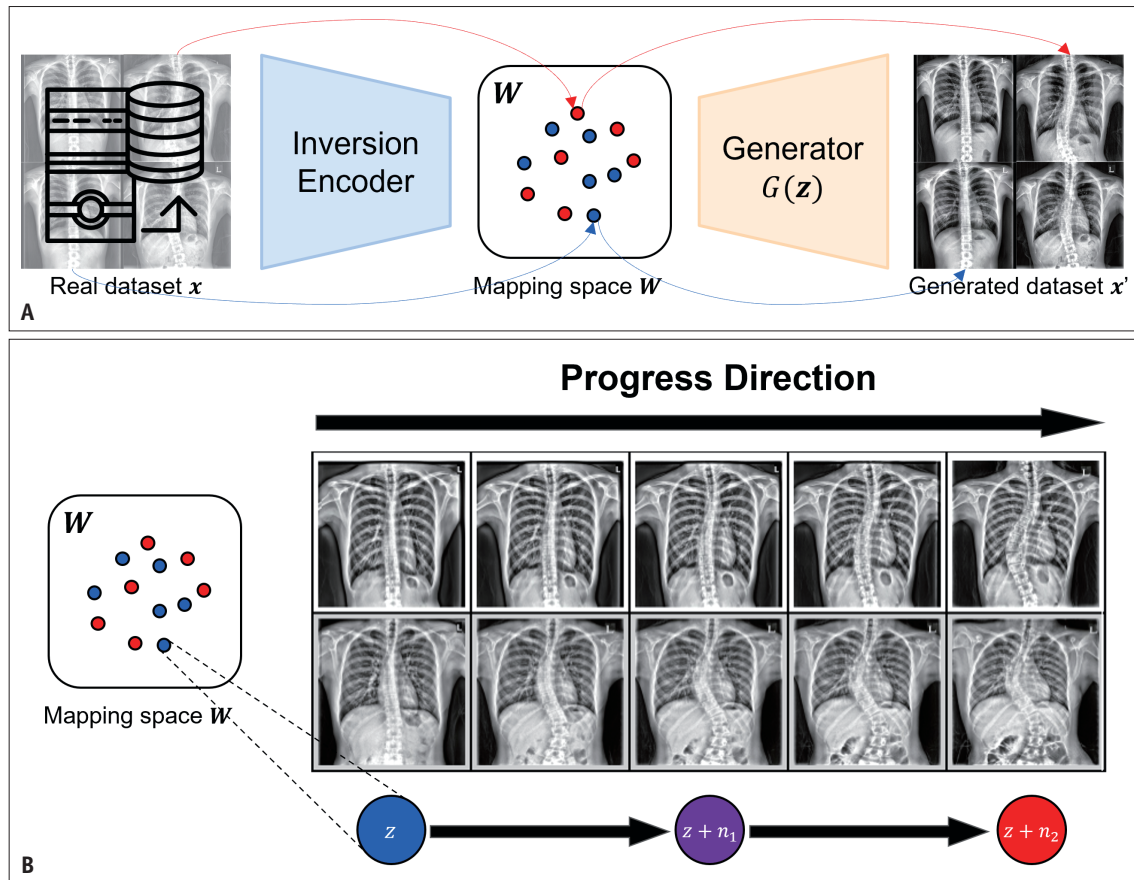
Generative AI models have demonstrated impressive capabilities in the field of medicine. However, despite their notable strengths, these models are associated with several limitations and potential drawbacks that warrant careful consideration during their application.

First, mode collapse, a phenomenon wherein the generator fails to provide a wide array of outputs, is a primary limitation of GANs that often results in repetitive samples or a limited set of variations [135-137]. This limitation influences the diversity and richness of the generated data.

Second, the production of high-quality and accurate samples that mirror the complexity of medical training data remains a challenge. The output may lack coherence or exhibit artifacts in some instances, which can undermine its reliability and usability in medical applications [31,138,139].

Third, these models tend to assimilate the biases present in the training data, which can lead to overfitting issues and a limited ability to generalize to new or unseen medical datasets. The assessment of the uncertainty or reliability of the generated medical samples is a significant challenge that hinders the quantification of predictive confidence and impacts their reliability in critical decision-making processes in healthcare.

Fourth, ethical concerns regarding the use of these models to generate false or misleading medical information, which can compromise authenticity and trustworthiness,



**Fig. 7.** Manipulating the inverted latent feature and editing images with the GAN inversion. GAN = generative adversarial network,  $x$  = dataset consists of real images,  $x'$  = dataset consists of generated images,  $W$  = latent space of dataset  $x$  mapped by the inversion encoder,  $z$  = latent feature of the real image,  $n_i$  = editing parameter of latent feature  $z$ ,  $z + n_i$  = edited latent feature  $z$

have arisen.

Fifth, the computational demands for training these models are substantial and require significant resources and time, thereby limiting their accessibility and scalability, particularly in resource-constrained healthcare environments.

Sixth, the interpretability and control of these models remain challenging, particularly in complex architectures such as GANs. The opaque nature of these models impedes their fine-tuning, understanding, and decision-making processes, which limits their practical applications in medical research and clinical settings [11]. Hallucinations, characterized by the generation of unrealistic or spurious outputs, challenge the credibility and reliability of generated content [140,141]. The outputs, while resembling authentic data, may contain elements or details that are absent in the training data, posing risks to medical decision-making and research.

Lastly, robust generalization of these generative models

to unseen medical data distributions remains a challenge that must be surmounted to enable their reliable and ethical use in medical research, diagnosis, and treatment planning.

Thus, the adoption of generative AI in clinical practice should be monitored by experts who can determine the reliability of the generated results. Furthermore, physicians and researchers in the medical field should be aware of the pitfalls and limitations of generative AI and exercise caution while implementing them. Guidelines for the use of generative AIs may facilitate a more comprehensive understanding of their advantages and pitfalls [142-144].

## CONCLUSION

The landscape of generative AI models encompasses various categories, including vision models (VAEs, GANs, and diffusion models) and language models (LLMs and VLMs). Each category comprises variants and applications tailored

for specific tasks, particularly medical imaging and analysis. LLMs, such as ChatGPT and GPT-4, built on the transformer architecture, excel in human-like text generation. In contrast, BERT, which is based on transformers, focuses on bidirectional language understanding. Recent advancements in VLMs have facilitated the execution of tasks, such as VQA, image recognition, and retrieval.

Generative models can significantly enhance image quality in clinical research, expedite MRI acquisition, generate synthetic data for rare diseases, and alleviate concerns regarding patient/data privacy in clinical AI research. Furthermore, the inversion of generative models facilitates the manipulation and editing of features within the latent space, thereby offering insights into latent representations, enabling control over data generation, and supporting various image manipulation tasks in the field of medicine.

Nevertheless, despite these benefits, generative AI is associated with multiple challenges, such as mode collapse, bias amplification, interpretability issues, computational demands, ethical concerns, and the potential to generate unrealistic data (hallucinations). Expert validation and awareness of these limitations play a crucial role in ensuring its responsible use in the medical context. Striking a balance between the potential benefits and disadvantages of generative AI is essential for the ethical and effective integration of these models into healthcare and research practices.

### Conflicts of Interest

Namkug Kim who is on the editorial board of the *Korean Journal of Radiology* was not involved in the editorial evaluation or decision to publish this article. All remaining authors have declared no conflicts of interest.

### Author Contributions

Conceptualization: Gil-Sun Hong, Namkug Kim. Funding acquisition: Gil-Sun Hong. Investigation: Kiduk Kim, Sunggu Kyung, Soyoung Lee, Kyungjin Cho, Ryoungwoo Jang, Sungwon Ham, Edward Choi. Project administration: Kiduk Kim. Supervision: Gil-Sun Hong, Namkug Kim. Visualization: Kiduk Kim, Gil-Sun Hong, Namkug Kim. Writing—original draft: Kiduk Kim, Sunggu Kyung, Soyoung Lee, Kyungjin Cho, Ryoungwoo Jang, Sungwon Ham, Edward Choi. Writing—review & editing: Kiduk Kim, Gil-Sun Hong, Namkug Kim.

### ORCID IDs

Kiduk Kim

<https://orcid.org/0000-0002-9659-897X>

Kyungjin Cho

<https://orcid.org/0000-0002-8111-5058>

Ryoungwoo Jang

<https://orcid.org/0000-0002-1511-7469>

Sunggu Kyung

<https://orcid.org/0000-0002-7582-9484>

Soyoung Lee

<https://orcid.org/0000-0003-1729-1775>

Sungwon Ham

<https://orcid.org/0000-0002-2886-6952>

Edward Choi

<https://orcid.org/0000-0002-5958-3509>

Gil-Sun Hong

<https://orcid.org/0000-0002-0068-9413>

Namkug Kim

<https://orcid.org/0000-0002-3438-2217>

### Funding Statement

This research was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI21C1148 and HI22C1723).

### REFERENCES

1. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets [accessed on August 28, 2023]. Available at: [https://papers.nips.cc/paper\\_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf](https://papers.nips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf)
2. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_CVPR\\_2019/papers/Karras\\_A\\_Style-Based\\_Generator\\_Architecture\\_for\\_Generative\\_Adversarial\\_Networks\\_CVPR\\_2019\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2019/papers/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.pdf)
3. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models [accessed on August 28, 2023]. Available at: <https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf>
4. Song J, Meng C, Ermon S. Denoising diffusion implicit models. arXiv [Preprint]. 2020 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2010.02502>
5. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S,

- Poole B. Score-based generative modeling through stochastic differential equations. arXiv [Preprint]. 2020 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2011.13456>
6. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training [accessed on August 28, 2023]. Available at: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)
  7. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners [accessed on August 28, 2023]. Available at: <https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf>
  8. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners [accessed on August 28, 2023]. Available at: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>
  9. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback [accessed on August 28, 2023]. Available at: [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf)
  10. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv [Preprint]. 2018 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1810.04805>
  11. Hong GS, Jang M, Kyung S, Cho K, Jeong J, Lee GY, et al. Overcoming the challenges in the development and implementation of artificial intelligence in radiology: a comprehensive review of solutions beyond supervised learning. *Korean J Radiol* 2023;24:1061-1080
  12. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv [Preprint]. 2013 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1312.6114>
  13. Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training generative adversarial networks with limited data [accessed on August 28, 2023]. Available at: <https://papers.nips.cc/paper/2020/file/8d30aa96e72440759f74bd2306c1fa3d-Paper.pdf>
  14. Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. PaLM: scaling language modeling with pathways. arXiv [Preprint]. 2022 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2204.02311>
  15. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. LLaMA: open and efficient foundation language models. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2302.13971>
  16. OpenAI. GPT-4 technical report. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2303.08774>
  17. Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, et al. Zero-shot text-to-image generation [accessed on November 28, 2023]. Available at: <https://proceedings.mlr.press/v139/ramesh21a.html>
  18. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents. arXiv [Preprint]. 2022 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2204.06125>
  19. Liu H, Li C, Wu Q, Lee YJ. Visual instruction tuning. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2304.08485>
  20. Kramer MA. Autoassociative neural networks. *Comput Chem Eng* 1992;16:313-328
  21. Sohn K, Lee H, Yan X. Learning structured output representation using deep conditional generative models [accessed on November 28, 2023]. Available at: [https://papers.nips.cc/paper\\_files/paper/2015/hash/8d55a249e6ba5c06772297520da2051-Abstract.html](https://papers.nips.cc/paper_files/paper/2015/hash/8d55a249e6ba5c06772297520da2051-Abstract.html)
  22. Ivanov O, Figurnov M, Vetrov D. Variational autoencoder with arbitrary conditioning. arXiv [Preprint]. 2018 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1806.02382>
  23. Chen RT, Li X, Grosse R, Duvenaud DK. Isolating sources of disentanglement in VAEs [accessed on November 28, 2023]. Available at: [https://proceedings.neurips.cc/paper\\_files/paper/2018/file/1ee3dfcd8a0645a25a35977997223d22-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2018/file/1ee3dfcd8a0645a25a35977997223d22-Paper.pdf)
  24. Vahdat A, Kautz J. NVAE: a deep hierarchical variational autoencoder [accessed on November 28, 2023]. Available at: <https://proceedings.neurips.cc/paper/2020/file/e3b21256183cf7c2c7a66be163579d37-Paper.pdf>
  25. Gregor K, Danihelka I, Graves A, Rezende D, Wierstra D. DRAW: a recurrent neural network for image generation [accessed on November 28, 2023]. Available at: <https://proceedings.mlr.press/v37/gregor15.pdf>
  26. Chung J, Kastner K, Dinh L, Goel K, Courville AC, Bengio Y. A recurrent latent variable model for sequential data [accessed on November 28, 2023]. Available at: <https://proceedings.neurips.cc/paper/2015/hash/b618c3210e934362ac261db280128c22-Abstract.html>
  27. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv [Preprint]. 2015 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1511.06434>
  28. Mirza M, Osindero S. Conditional generative adversarial nets. arXiv [Preprint]. 2014 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1411.1784>
  29. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. arXiv [Preprint]. 2017 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1710.10196>
  30. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of StyleGAN [accessed on August 28, 2023]. Available at: <https://openaccess.thecvf.com>

- com/content\_CVPR\_2020/papers/Karras\_Analyzing\_and\_Improving\_the\_Image\_Quality\_of\_StyleGAN\_CVPR\_2020\_paper.pdf
31. Karras T, Aittala M, Laine S, Härkönen E, Hellsten J, Lehtinen J, et al. Alias-free generative adversarial networks [accessed on August 28, 2023]. Available at: [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/076ccd93ad68be51f23707988e934906-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/076ccd93ad68be51f23707988e934906-Paper.pdf)
  32. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_ICCV\\_2017/papers/Zhu\\_Unpaired\\_Image-To-Image\\_Translation\\_ICCV\\_2017\\_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.pdf)
  33. Gravina M, Marrone S, Docimo L, Santini M, Fiorelli A, Parmeggiani D, et al. Leveraging CycleGAN in lung CT sinogram-free kernel conversion. Proceedings of the 21st International Conference on Image Analysis and Processing-ICIAP 2022; 2022 May 23-27; Lecce, Italy; Cham: Springer International Publishing, 2022:100-110
  34. Yang S, Kim EY, Ye JC. Continuous conversion of CT kernel using switchable CycleGAN with AdaIN. *IEEE Trans Med Imaging* 2021;40:3015-3029
  35. Tang C, Li J, Wang L, Li Z, Jiang L, Cai A, et al. Unpaired low-dose CT denoising network based on cycle-consistent generative adversarial network with prior image information. *Comput Math Methods Med* 2019;2019:8639825
  36. Kwon T, Ye JC. Cycle-free CycleGAN using invertible generator for unsupervised low-dose ct denoising. *IEEE Trans Comput Imaging* 2021;7:1354-1368
  37. Gu J, Ye JC. AdaIN-based tunable CycleGAN for efficient unsupervised low-dose CT denoising. *IEEE Trans Comput Imaging* 2021;7:73-85
  38. Yan C, Lin J, Li H, Xu J, Zhang T, Chen H, et al. Cycle-consistent generative adversarial network: effect on radiation dose reduction and image quality improvement in ultralow-dose CT for evaluation of pulmonary tuberculosis. *Korean J Radiol* 2021;22:983-993
  39. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Choi\\_StarGAN\\_Unified\\_Generative\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Choi_StarGAN_Unified_Generative_CVPR_2018_paper.html)
  40. Sohail M, Riaz MN, Wu J, Long C, Li S. Unpaired multi-contrast MR image synthesis using generative adversarial networks. Proceedings of the 4th International Workshop on Simulation and Synthesis in Medical Imaging-SASHIMI 2019; 2019 Oct 13; Shenzhen, China; Cham: Springer International Publishing, 2019:22-31
  41. Liao Z, Jafari MH, Girgis H, Gin K, Rohling R, Abolmaesumi P, et al. Echocardiography view classification using quality transfer star generative adversarial networks. Proceedings of 22nd International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2019; 2019 Oct 13-17; Shenzhen, China; Cham: Springer International Publishing, 2019:687-695
  42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need [accessed on August 28, 2023]. Available at: [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html)
  43. Yang Z, Xu X, Yao B, Zhang S, Rogers E, Intille S, et al. Talk2Care: facilitating asynchronous patient-provider communication with large-language-model. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2309.09357>
  44. Kahambing JG. ChatGPT, public health communication and 'intelligent patient companionship.' *J Public Health (Oxf)* 2023;45:e590
  45. Jeblick K, Schachtner B, Dext J, Mittermeier A, Stüber AT, Topalis J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol* 2023 Oct 5 [Epub]. <https://doi.org/10.1007/s00330-023-10213-1>
  46. Uppal S, Bhagat S, Hazarika D, Majumder N, Poria S, Zimmermann R, et al. Multimodal research in vision and language: a review of current and emerging trends. *Inf Fusion* 2022;77:149-171
  47. Antol S, Agrawal A, Lu J, Mitchell M, Batra D, Zitnick CL, et al. VQA: visual question answering [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_iccv\\_2015/html/Antol\\_VQA\\_Visual\\_Question\\_ICCV\\_2015\\_paper.html](https://openaccess.thecvf.com/content_iccv_2015/html/Antol_VQA_Visual_Question_ICCV_2015_paper.html)
  48. Zellers R, Bisk Y, Farhadi A, Choi Y. From recognition to cognition: visual commonsense reasoning [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_CVPR\\_2019/papers/Zellers\\_From\\_Recognition\\_to\\_Cognition\\_Visual\\_Commonsense\\_Reasoning\\_CVPR\\_2019\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2019/papers/Zellers_From_Recognition_to_Cognition_Visual_Commonsense_Reasoning_CVPR_2019_paper.pdf)
  49. Hossain MZ, Sohel F, Shiratuddin MF, Laga H. A comprehensive survey of deep learning for image captioning. *ACM Comput Surv* 2019;51:1-36
  50. de Rosa GH, Papa JP. A survey on text generation using generative adversarial networks. *Pattern Recognit* 2021;119:108098
  51. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision [accessed on August 28, 2023]. Available at: <https://proceedings.mlr.press/v139/radford21a/radford21a.pdf>
  52. Jia C, Yang Y, Xia Y, Chen YT, Parekh Z, Pham H, et al. Scaling up visual and vision-language representation learning with noisy text supervision [accessed on August 28, 2023]. Available at: <http://proceedings.mlr.press/v139/jia21b.html>
  53. Chen X, Ma L, Chen J, Jie Z, Liu W, Luo J. Real-time referring expression comprehension by single-stage grounding network. arXiv [Preprint]. 2018 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/>

- arXiv.1812.03426
54. Nagaraja VK, Morariu VI, Davis LS. Modeling context between objects for referring expression understanding. Proceedings of 14th European Conference on Computer Vision-ECCV 2016; 2016 October 11-14; Amsterdam, The Netherlands; Cham: Springer International Publishing, 2016:792-807
  55. Cao M, Li S, Li J, Nie L, Zhang M. Image-text retrieval: a survey on recent research and development. arXiv [Preprint]. 2022 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2203.14713>
  56. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2304.02643>
  57. Zhang K, Liu D. Customized segment anything model for medical image segmentation. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2304.13785>
  58. Zhang Y, Jiao R. Towards segment anything model (SAM) for medical image segmentation: a survey. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2305.03678>
  59. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nature Communications* 2024;15:654
  60. Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. Segment anything model for medical image analysis: an experimental study. *Med Image Anal* 2023;89:102918
  61. Hu C, Li X. When SAM meets medical images: an investigation of segment anything model (SAM) on multi-phase liver tumor segmentation. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2304.08506>
  62. Jung KH. Uncover this tech term: foundation model. *Korean J Radiol* 2023;24:1038-1041
  63. Alayrac JB, Donahue J, Luc P, Miech A, Barr I, Hasson Y, et al. Flamingo: a visual language model for few-shot learning [accessed on August 28, 2023]. Available at: [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/960a172bc77bf0177ccccbb411a7d800-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/960a172bc77bf0177ccccbb411a7d800-Paper-Conference.pdf)
  64. Yuan L, Chen D, Chen YL, Codella N, Dai X, Gao J, et al. Florence: a new foundation model for computer vision. arXiv [Preprint]. 2021 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2111.11432>
  65. Li C, Wong C, Zhang S, Usuyama N, Liu H, Yang J, et al. LLaVA-Med: training a large language-and-vision assistant for biomedicine in one day. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2306.00890>
  66. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature* 2023;620:172-180
  67. Ghesu FC, Georgescu B, Mansoor A, Yoo Y, Neumann D, Patel P, et al. Self-supervised learning from 100 million medical images. arXiv [Preprint]. 2022 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2201.01283>
  68. Cho K, Kim KD, Nam Y, Jeong J, Kim J, Choi C, et al. CheSS: chest X-ray pre-trained model via self-supervised contrastive learning. *J Digit Imaging* 2023;36:902-910
  69. Wu C, Zhang X, Zhang Y, Wang Y, Xie W. Towards generalist foundation model for radiology by leveraging web-scale 2D&3D medical data. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2308.02463>
  70. Kang E, Koo HJ, Yang DH, Seo JB, Ye JC. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Med Phys* 2019;46:550-562
  71. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 2017;36:2536-2545
  72. Wang J, Zhao Y, Noble JH, Dawant BM. Conditional generative adversarial networks for metal artifact reduction in CT images of the ear. Proceedings of 21st International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2018; 2018 Sep 16-20; Granada, Spain; Cham: Springer International Publishing, 2018:3-11
  73. Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. *Phys Med Biol* 2019;64:125002
  74. Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys* 2019;46:3998-4009
  75. Kim KH, Do WJ, Park SH. Improving resolution of MR images with an adversarial network incorporating images with different contrast. *Med Phys* 2018;45:3120-3131
  76. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging* 2018;37:1488-1497
  77. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging* 2018;37:1310-1321
  78. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans Biomed Eng* 2018;65:2720-2730
  79. Yao Z, Luo T, Dong Y, Jia X, Deng Y, Wu G, et al. Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis. *Nat Commun* 2023;14:788
  80. Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol* 2018;63:185001
  81. Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys* 2019;46:3565-3581



82. Conte GM, Weston AD, Vogelsang DC, Philbrick KA, Cai JC, Barbera M, et al. Generative adversarial networks to synthesize missing T1 and FLAIR MRI sequences for use in a multisequence brain tumor segmentation model. *Radiology* 2021;299:313-323
83. Lei Y, Dong X, Tian Z, Liu Y, Tian S, Wang T, et al. CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. *Med Phys* 2020;47:530-540
84. Chung M, Kong ST, Park B, Chung Y, Jung KH, Seo JB. Utilizing synthetic nodules for improving nodule detection in chest radiographs. *J Digit Imaging* 2022;35:1061-1068
85. Al Khalil Y, Amirrajab S, Lorenz C, Weese J, Pluim J, Breeuwer M. On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images. *Med Image Anal* 2023;84:102688
86. Jayachandran Preetha C, Meredig H, Brugnara G, Mahmutoglu MA, Foltyn M, Isensee F, et al. Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. *Lancet Digit Health* 2021;3:e784-e794
87. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019;9:16884
88. Park JE, Vollmuth P, Kim N, Kim HS. Research highlight: use of generative images created with artificial intelligence for brain tumor imaging. *Korean J Radiol* 2022;23:500-504
89. Bae K, Oh DY, Yun ID, Jeon KN. Bone suppression on chest radiographs for pulmonary nodule detection: comparison between a generative adversarial network and dual-energy subtraction. *Korean J Radiol* 2022;23:139-149
90. Tanner C, Ozdemir F, Profanter R, Vishnevsky V, Konukoglu E, Goksel O. Generative adversarial networks for MR-CT deformable image registration. arXiv [Preprint]. 2018 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.1807.07349>
91. Yan P, Xu S, Rastinehad AR, Wood BJ. Adversarial image registration with application for MR and TRUS image fusion. Proceedings of 9th International Workshop on Machine Learning in Medical Imaging-MLMI 2018; 2018 Sep 16; Granada, Spain; Cham: Springer International Publishing, 2018:197-204
92. Wolleb J, Sandkühler R, Cattin PC. DeScarGAN: disease-specific anomaly detection with weak supervision. Proceedings of 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2020; 2020 Oct 4-8; Lima, Peru; Cham: Springer International Publishing, 2020:14-24
93. Nakao T, Hanaoka S, Nomura Y, Murata M, Takenaga T, Miki S, et al. Unsupervised deep anomaly detection in chest radiographs. *J Digit Imaging* 2021;34:418-427
94. Lee S, Jeong B, Kim M, Jang R, Paik W, Kang J, et al. Emergency triage of brain computed tomography via anomaly detection with a deep generative model. *Nat Commun* 2022;13:4251
95. van Hespen KM, Zwanenburg JJM, Dankbaar JW, Geerlings MI, Hendrikse J, Kuijf HJ. An anomaly detection approach to identify chronic brain infarcts on MRI. *Sci Rep* 2021;11:7714
96. Khosla M, Jamison K, Kuceyeski A, Sabuncu MR. Detecting abnormalities in resting-state dynamics: an unsupervised learning approach. Proceedings of 10th International Workshop on Machine Learning in Medical Imaging; 2019 Oct 13; Shenzhen, China; Cham: Springer International Publishing, 2019:301-309
97. Han C, Rundo L, Murao K, Noguchi T, Shimahara Y, Milacski ZÁ, et al. MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. *BMC Bioinformatics* 2021;22(Suppl 2):31
98. Bowles C, Gunn R, Hammers A, Rueckert D. Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks [accessed on August 28, 2023]. Available at: <https://doi.org/10.1117/12.2293256>
99. Liu X, Xie Y, Cheng J, Diao S, Tan S, Liang X. Diffusion probabilistic priors for zero-shot low-dose CT image denoising. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2305.15887>
100. Gao Q, Li Z, Zhang J, Zhang Y, Shan H. CoreDiff: contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.1109/TMI.2023.3320812>
101. Li Q, Li C, Yan C, Li X, Li H, Zhang T, et al. Ultra-low dose CT image denoising based on conditional denoising diffusion probabilistic model [accessed on August 28, 2023]. Available at: <https://doi.org/10.1109/CyberC55534.2022.00041>
102. Gao Q, Shan H. CoCoDiff: a contextual conditional diffusion model for low-dose CT image denoising [accessed on August 28, 2023]. Available at: <https://doi.org/10.1117/12.2634939>
103. Selim M, Zhang J, Brooks MA, Wang G, Chen J. DiffusionCT: latent diffusion model for CT image standardization. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2301.08815>
104. Chung H, Ye JC. Score-based diffusion models for accelerated MRI. *Med Image Anal* 2022;80:102479
105. Xia W, Lyu Q, Wang G. Low-dose CT using denoising diffusion probabilistic model for 20x speedup. arXiv [Preprint]. 2022 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2209.15136>
106. Huang J, Aviles-Rivero AI, Schönlieb CB, Yang G. CDiffMR: can we replace the gaussian noise with K-space undersampling for fast MRI? Proceedings of 26th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2023; 2023 Oct 8-12; Vancouver, Canada; Cham: Springer International Publishing, 2023:3-12
107. Pan S, Abouei E, Wynne J, Wang T, Qiu RL, Li Y, et al. Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model. arXiv [Preprint]. 2023

- [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2305.19467>
108. Lyu Q, Wang G. Conversion between CT and MRI images using diffusion and score-matching models. arXiv [Preprint]. 2022 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2209.12104>
  109. Özbey M, Dalmaz O, Dar SU, Bedel HA, Ş Ö, Güngör A, et al. Unsupervised medical image translation with adversarial diffusion models [accessed on August 28, 2023]. Available at: <https://doi.org/10.1109/TMI.2023.3290149>
  110. Wu J, Fu R, Fang H, Zhang Y, Yang Y, Xiong H, et al. MedSegDiff: medical image segmentation with diffusion probabilistic model. arXiv [Preprint]. 2022 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2211.00611>
  111. Wu J, Fu R, Fang H, Zhang Y, Xu Y. MedSegDiff-V2: diffusion based medical image segmentation with transformer. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2301.11798>
  112. Wolleb J, Sandkühler R, Bieder F, Valmaggia P, Cattin PC. Diffusion models for implicit image segmentation ensembles [accessed on August 28, 2023]. Available at: <https://proceedings.mlr.press/v172/wolleb22a.html>
  113. Rahman A, Valanarasu JMJ, Hacıhaliloglu I, Patel VM. Ambiguous medical image segmentation using diffusion models [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content/CVPR2023/html/Rahman\\_Ambiguous\\_Medical\\_Image\\_Segmentation\\_Using\\_Diffusion\\_Models\\_CVPR\\_2023\\_paper.html](https://openaccess.thecvf.com/content/CVPR2023/html/Rahman_Ambiguous_Medical_Image_Segmentation_Using_Diffusion_Models_CVPR_2023_paper.html)
  114. Kim B, Han I, Ye JC. DiffuseMorph: unsupervised deformable image registration using diffusion model. Proceedings of 17th European Conference on Computer Vision-ECCV 2022; 2022 Oct 23-27; Tel Aviv, Israel; Cham: Springer International Publishing, 2022:347-364
  115. Fontanella A, Mair G, Wardlaw J, Trucco E, Storkey A. Diffusion models for counterfactual generation and anomaly detection in brain images. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2308.02062>
  116. Wolleb J, Bieder F, Sandkühler R, Cattin PC. Diffusion models for medical anomaly detection. Proceedings of 25th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2022; 2022 Sep 18-22; Singapore; Cham: Springer International Publishing, 2022:35-45
  117. Li J, Cao H, Wang J, Liu F, Dou Q, Chen G, et al. Fast non-markovian diffusion model for weakly supervised anomaly detection in brain MR images. Proceedings of 26th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2023; 2023 Oct 8-12; Vancouver, Canada; Cham: Springer International Publishing, 2023:579-589
  118. Pinaya WH, Graham MS, Gray R, Da Costa PF, Tudosiu PD, Wright P, et al. Fast unsupervised brain anomaly detection and segmentation with diffusion models. Proceedings of 25th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2022; 2022 Sep 18-22; Singapore; Cham: Springer International Publishing, 2022:705-714
  119. Behrendt F, Bhattacharya D, Krüger J, Opfer R, Schlaefer A. Patched diffusion models for unsupervised anomaly detection in brain MRI. arXiv [Preprint]. 2023 [accessed on August 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2303.03758>
  120. Xia W, Zhang Y, Yang Y, Xue JH, Zhou B, Yang MH. GAN inversion: a survey [accessed on August 28, 2023]. Available at: <https://doi.org/10.1109/TPAMI.2022.3181070>
  121. Mokady R, Hertz A, Aberman K, Pritch Y, Cohen-Or D. Null-text inversion for editing real images using guided diffusion models [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content/CVPR2023/html/Mokady\\_NULL-Text\\_Inversion\\_for\\_Editing\\_Real\\_Images\\_Using\\_Guided\\_Diffusion\\_Models\\_CVPR\\_2023\\_paper.html](https://openaccess.thecvf.com/content/CVPR2023/html/Mokady_NULL-Text_Inversion_for_Editing_Real_Images_Using_Guided_Diffusion_Models_CVPR_2023_paper.html)
  122. Zhu J, Shen Y, Zhao D, Zhou B. In-domain GAN inversion for real image editing. In: Vedaldi A, Bischof H, Brox T, Frahm JM. *Computer vision – ECCV 2020*. Cham: Springer, 2020:592-608
  123. Wang T, Zhang Y, Fan Y, Wang J, Chen Q. High-fidelity GAN inversion for image attribute editing [accessed on August 28, 2023]. Available at: [https://openaccess.thecvf.com/content/CVPR2022/papers/Wang\\_High-Fidelity\\_GAN\\_Inversion\\_for\\_Image\\_Attribute\\_Editing\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Wang_High-Fidelity_GAN_Inversion_for_Image_Attribute_Editing_CVPR_2022_paper.pdf)
  124. Ren Z, Yu SX, Whitney D. Controllable medical image generation via GAN. *J Percept Imaging* 2022;5:000502-1-000502-15
  125. Fetty L, Bylund M, Kuess P, Heilemann G, Nyholm T, Georg D, et al. Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Z Med Phys* 2020;30:305-314
  126. Lee JS, Shin K, Ryu SM, Jegal SG, Lee W, Yoon MA, et al. Screening of adolescent idiopathic scoliosis using generative adversarial network (GAN) inversion method in chest radiographs. *PLoS One* 2023;18:e0285489
  127. Marcel S, Millán Jdel R. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans Pattern Anal Mach Intell* 2007;29:743-752
  128. Topol EJ. What's lurking in your electrocardiogram? *Lancet* 2021;397:785
  129. Arora A. Synthetic data: the future of open-access health-care datasets? *Lancet* 2023;401:997
  130. Elreedy D, Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Inf Sci* 2019;505:32-64
  131. van Breugel B, Kyono T, Berrevoets J, van der Schaar M. DECAF: generating fair synthetic data using causally-aware generative networks [accessed on November 28, 2023]. Available at: <https://proceedings.neurips.cc/paper/2021/hash/ba9fab001f67381e56e410575874d967-Abstract.html>
  132. Rajotte JF, Bergen R, Buckeridge DL, El Emam K, Ng R, Strome E. Synthetic data as an enabler for machine learning applications

- in medicine. *iScience* 2022;25:105331
133. Banerjee R, Midha S, Kelkar AH, Goodman A, Prasad V, Mohyuddin GR. Synthetic control arms in studies of multiple myeloma and diffuse large B-cell lymphoma. *Br J Haematol* 2022;196:1274-1277
  134. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials - a primer for researchers. *Clin Epidemiol* 2020;12:457-467
  135. Thanh-Tung H, Tran T. Catastrophic forgetting and mode collapse in GANs [accessed on November 28, 2023]. Available at: <https://doi.org/10.1109/IJCNN48605.2020.9207181>
  136. Durall R, Chatzimichailidis A, Labus P, Keuper J. Combating mode collapse in GAN training: an empirical analysis using hessian eigenvalues. arXiv [Preprint]. 2020 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2012.09673>
  137. Bau D, Zhu JY, Wulff J, Peebles W, Strobelt H, Zhou B, et al. Seeing what a GAN cannot generate [accessed on November 28, 2023]. Available at: [https://openaccess.thecvf.com/content\\_ICCV\\_2019/html/Bau\\_Seeing\\_What\\_a\\_GAN\\_Cannot\\_Generate\\_ICCV\\_2019\\_paper.html](https://openaccess.thecvf.com/content_ICCV_2019/html/Bau_Seeing_What_a_GAN_Cannot_Generate_ICCV_2019_paper.html)
  138. Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts. *Distill* 2016;1:e3
  139. Yin Y, Huang L, Liu Y, Huang K. DiffGAR: model-agnostic restoration from generative artifacts using image-to-image diffusion models [accessed on November 28, 2023]. Available at: <https://dl.acm.org/doi/abs/10.1145/3577530.3577539>
  140. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv* 2023;55:1-38
  141. Mündler N, He J, Jenko S, Vechev M. Self-contradictory hallucinations of large language models: evaluation, detection and mitigation. arXiv [Preprint]. 2023 [accessed on November 28, 2023]. Available at: <https://doi.org/10.48550/arXiv.2305.15852>
  142. Koga S. The integration of large language models such as ChatGPT in scientific writing: harnessing potential and addressing pitfalls. *Korean J Radiol* 2023;24:924-925
  143. Hwang SI, Lim JS, Lee RW, Matsui Y, Iguchi T, Hiraki T, et al. Is ChatGPT a “fire of prometheus” for non-native english-speaking researchers in academic writing? *Korean J Radiol* 2023;24:952-959
  144. Park SH. Use of generative artificial intelligence, including large language models such as ChatGPT, in scientific publications: policies of KJR and prominent authorities. *Korean J Radiol* 2023;24:715-718