

# Exploring Foundation Models for Synthetic Medical Imaging: A Study on Chest X-Rays and Fine-Tuning Techniques

Davide Clode da Silva, Marina Musse Bernardes, Nathália Giacomini Ceretta, Gabriel Vaz de Souza, Gabriel Fonseca Silva, Rafael Heitor Bordini and Soraia Raupp Musse  
Pontifical Catholic University of Rio Grande do Sul (PUCRS), RS, Brazil  
Email: gabriel.fonseca94@edu.pucrs.br, soraia.musse@pucrs.br

**Abstract**—Machine learning has significantly advanced healthcare by aiding in disease prevention and treatment identification. However, accessing patient data can be challenging due to privacy concerns and strict regulations. Generating synthetic, realistic data offers a potential solution for overcoming these limitations, and recent studies suggest that fine-tuning foundation models can produce such data effectively. In this study, we explore the potential of foundation models for generating realistic medical images, particularly chest x-rays, and assess how their performance improves with fine-tuning. We propose using a Latent Diffusion Model, starting with a pre-trained foundation model and refining it through various configurations. Additionally, we performed experiments with input from a medical professional to assess the realism of the images produced by each trained model.

## I. INTRODUCTION

In recent years, Machine Learning (ML) has played a crucial role in healthcare. For instance, in disease prevention and treatment, ML can help analyze large datasets to identify trends and predict disease outcomes, such as modeling the progression and treatment of cancerous conditions [1]. Nevertheless, the adoption of ML techniques in healthcare has been slow due to factors like scarcity of patient data, data privacy concerns, regulatory requirements, and the critical nature of healthcare decisions [2], [3].

One of the main barriers to accessing patient medical data is the need to protect sensitive and confidential information from unauthorized access. Additionally, the lack of standardized records and the effort required to collect medical data pose significant challenges [4], [5]. Generating synthetic, realistic, and high-quality medical data could be a viable alternative to mitigate some of these issues. The industry predicts a significant increase in the availability of synthetic data in the coming years, potentially doubling the amount of real data currently available [6]. Examples include generative models for creating photorealistic images from natural language descriptions [7] and improving GAN-based (Generative Adversarial Network) models' performance on super-resolution images [8].

In the context of medical images, generative models may offer a practical solution, with some work focusing on fine-tuning foundation models using small datasets [9]. Foundation Models (FMs) are machine learning models trained on a

wide range of generalized and unlabeled data, typically using self-supervision techniques. Examples of Foundation Models include ELMo [10], GPT-3 [11], CLIP [12], ResNet [13], DALL-E [14], and Stable Diffusion [15]. These models have achieved significant advancements in various complex tasks [16], such as Question Answering [11], Knowledge Base Construction [17], and Information Retrieval [18]. Fine-tuning such models targets the foundation model generalization capabilities into specific applications.

One example is the MedCLIP project [19], where a Contrastive Language-Image Pre-Training (CLIP) model was adapted using the ROCO dataset [12], which includes fine-tuned images and corresponding captions. Researchers claim that the adjusted model, MedCLIP, can identify higher-level characteristics such as the image modality, distinguishing between PET (Positron Emission Tomography) scans and ultrasound scans. Other studies have used generative model approaches for various purposes, including synthesizing realistic medical data [20], [21] and augmenting datasets for deep learning model training [8].

In this work, we present an initial exploration of the capabilities of foundational models on generating realistic medical images and how their performance is impacted by fine-tuning. We use a small dataset during the fine-tuning process to assess the ability of FMs to learn effectively with limited data. We focus on generating chest x-ray images, considering healthy and unhealthy diagnoses. We propose utilizing a Latent Diffusion Model (LDM) approach, employing a pre-trained foundation model as a basis and later fine-tuning it with different configurations. Furthermore, we conducted experiments with the assistance of a medical professional to evaluate the realism of images generated by each model trained.

The remainder of this work is organized as follows: Section II presents related work that has played an important role in understanding different approaches to generating synthetic medical data. Section III details our proposed method for generating synthetic medical images. Section IV presents the experiments conducted and preliminary results achieved. Finally, Section V presents our final considerations, limitations, and possible future contributions.

## II. RELATED WORK

The literature presents various image-generated techniques, often based on a textual description. Chen et al [22] proposed Re-Imagen (Retrieval-Augmented Text-to-Image Generator), a generative model that utilizes retrieved information to produce accurate images. It is distinguished by its capability to generate faithful images, even for rare or never-before-seen entities. Zhou et al. [23] introduced LAFITE (LAnguage-Free training for Text-to-image gEneration) to address one of the major challenges in training text-to-image generation models: the requirement for a large number of image-text pairs. The authors claim that their proposed model can be trained without any text data, and it has shown promising results.

Focusing on medical images, Pinaya et al. [21] proposed a model to generate synthetic images from high-resolution brain MRIs. Their model uses data to learn the probabilistic distribution of brain images based on covariates like age, sex, and brain structure volumes. The authors employed Latent Diffusion Models (LDM), which combine autoencoders to compress input data into a lower-dimensional latent representation with the generative modeling properties of diffusion models. The compression model, crucial for enabling the scalability of high-resolution medical images, was trained using a combination of perceptual loss and an adversarial objective based on patches—specific small areas of an image used to modify or manipulate parts of the image to create adversaries or deceive machine learning algorithms.

Ali et al. [24] investigated medical image synthesis using diffusion models. Initially, the authors used a pre-trained DALLE2 [14] model to generate lung x-ray and CT images from text prompts, then, they trained a stable diffusion model on 3,165 x-ray images. For evaluation, two independent radiologists conducted a qualitative analysis by labeling randomly chosen samples as real, fake, or uncertain. The results indicated that the diffusion model could effectively translate features specific to certain medical conditions into chest X-rays or CT images.

Packhauser et al. [25] utilized (LDM) to generate high-quality chest x-ray images while preserving the privacy of sensitive biometric information. Conditional information was embedded using a trainable lookup table combined with cross-attention at the U-Net bottleneck. The model was trained on a dataset of chest X-rays from 30,805 patients, including metadata with 14 abnormality labels and an additional class for healthy individuals. The generated dataset was evaluated in a thoracic abnormality classification task, and the approach outperformed GAN-based methods.

Other recent work focuses on image generation for dataset augmentation. Sundaram et al. [6] proposed using Generative Adversarial Networks (GANs) to augment a chest x-ray dataset to address class imbalance. Their strategy involved creating synthetic chest X-ray images featuring at least one of three underrepresented pathologies: lung injury, pleural injury, or fracture. These synthetic images were added to the original dataset to reduce class imbalance. Motamed et al.

[20] introduced the Inception Augmentation GAN (IAGAN) for chest x-ray data augmentation, targeting semi-supervised detection of pneumonia and COVID-19. Inspired by the Data Augmentation Generative Adversarial Networks (DAGAN) [26], their model aimed to generate synthetic data to enhance training datasets for other models.

## III. PROPOSED MODEL

In this work, our main task is to perform the fine-tuning of a Latent Diffusion Model in order to generate high-resolution synthetic chest x-ray images. Section III-A describe the dataset utilized for fine-tuning, composed of real chest x-ray images. Section III-B describes the fine-tuning process, resources and parameterization utilized.

### A. Dataset

In this study, we use the Montgomery County CXR Set dataset for tuberculosis, developed by the National Library of Medicine in collaboration with the Department of Health and Human Services in Montgomery<sup>1</sup>. The dataset is publicly available<sup>2</sup> and is composed of 138 postero-anterior chest x-ray images. Of these, 80 are from normal (or healthy) cases, and 58 are from abnormal (or unhealthy) cases with manifestations consistent with tuberculosis.

Additionally, the dataset includes consensus annotations from two radiologists for resized 1024 × 1024 images and a report describing the imaging results [27]. Figure 1 presents sample images of both cases and their respective annotations. All images are anonymized and available in PNG format. Based on the 138 x-rays available, we used a subset consisting of 30 images (50% healthy images and 50% unhealthy images) to fine-tune the models. We chose to work with a data set of this size as an initial exploration so that we could obtain more precise guidance on the next steps to be taken in future work.

### B. Fine-tuning

In our approach, we utilized the Kohya-ss GUI<sup>3</sup>, a friendly user interface that allows the setup, training and fine-tuning of diffusion models. The interface provides different techniques for fine-tuning, such as DreamBooth [28] and LoRA (Low-Rank Adaptation) [29]. We believe LoRA to be a more appropriate option for our method, given its training efficiency by having a smaller number of parameters, reduced hardware requirements for adaptive optimizers, versatility in fine-tuning, and a smaller final model [29]. Additionally, we utilized TensorBoard<sup>4</sup> alongside the Kohya-ss GUI as a visualization toolkit for the loss and learning rates during the fine-tuning.

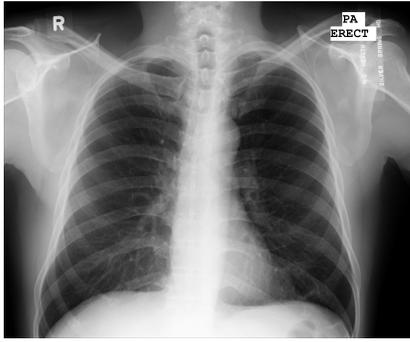
We selected “*stable-diffusion-v1-5*”, proposed by Rombach et al. [15], as our foundation model, which is capable of generating realistic images based on textual descriptions. We

<sup>1</sup>Information on the research is available at: <https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html>

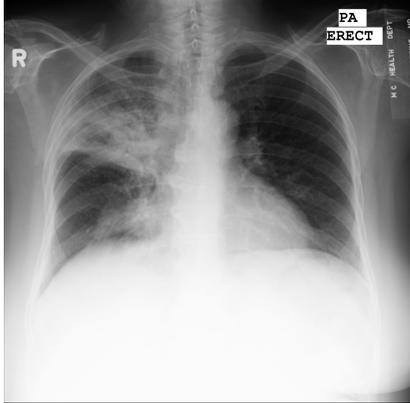
<sup>2</sup>The dataset is available for download at: <https://data.lhncbc.nlm.nih.gov/public/Tuberculosis-Chest-X-ray-Datasets/Shenzhen-Hospital-CXR-Set/index.html>

<sup>3</sup>Available at: [https://github.com/bmaltais/kohya\\_ss](https://github.com/bmaltais/kohya_ss)

<sup>4</sup>Available at: <https://github.com/tensorflow/tensorboard>



(a) “Normal chest x-ray”



(b) “Large infiltrate Right Upper Lobe with cavitation plus infiltrate in RML. Consistent with active cavitory TB.”

Fig. 1. Sample images from the dataset and their respective annotations [27]. (a) presents a normal case, and (b) an abnormal case.

performed a fine-tuning on this foundation model utilizing the chest x-ray dataset described in Section III-A. For the fine-tuning process, we consider different optimizers, described next:

- AdamW8bit: is a variant of Adam, a stochastic optimization method designed for large-scale machine learning problems. It adjusts the learning rate for each model weight individually and computes adaptive learning rates for different parameters [30]–[32].
- Adafactor: aims to overcome the memory requirements of stochastic optimization methods like RMSProp and Adam by storing only the row and column sums of these averages, reducing memory usage while retaining adaptability [33].
- DAdaptSGD: is a variant of DAdaption, specifically designed to automatically determine the learning rate in the Stochastic Gradient Descent (SGD) optimization algorithm [34].
- Prodigy: is an optimizer that estimates the distance to the solution, allowing for optimal learning rate adjustment in adaptive methods like Adagrad and Adam. It’s an adaptation of the D-Adaptation method for learning-rate-free learning, where the learning rate is set automatically [35].

We fine-tuned five distinct models, two utilizing the AdamW8bit, and one with each remaining optimizer. Table I presents the parameterization of each training. All models were trained for 100 epochs. Finally, we utilized the Stable Diffusion Web UI<sup>5</sup> for generating the images. The interface allows the user to select different models, image quantities and dimensions, seeds, and the prompt used for the generation.

TABLE I

PARAMETRIZATION OF THE FINE-TUNING PROCESS. ALL MODELS WERE TRAINED USING “stable-diffusion-v1-5” AS FOUNDATION MODEL. THE ADAM8BIT OPTIMIZER WAS USED IN TWO DISTINCT MODELS, M1 USING ITS DEFAULT PARAMETERS, AND M2 USING A CONFIGURATION SIMILAR TO THE REMAINING OPTIMIZERS.

Model	M1	M2	M3	M4	M5
Optimizer	AdamW8bit	AdamW8bit	Adafactor	DAdaptSGD	Prodigy
Clip Skip	2	2	2	2	2
Epochs	100	100	100	100	100
LR	$1.10^{-4}$	$1.10^{-4}$	$1.10^{-4}$	$1.10^{-4}$	$1.10^{-4}$
Max Resolution	512x512	512x512	512x512	512x512	512x512
LR Scheduler	constant	constant	constant	constant	constant
Train Batch Size	2	2	2	2	2
Text Encoder LR	$5.10^{-5}$	$5.10^{-5}$	$5.10^{-5}$	$1.10^{-5}$	$1.10^{-5}$
Unet lr	$1.10^{-4}$	$1.10^{-4}$	$1.10^{-4}$	$1.10^{-5}$	$1.10^{-5}$
VAE Batch Size	0	32	32	32	32
Noise Offset Type	Original	Multires	Multires	Multires	Multires
Noise Discount	0	0.1	0.1	0.1	0.1
Noise Iteration	0	6	6	6	6

#### IV. EXPERIMENTAL RESULTS

For our experiments, we generated images of chest x-rays using six different models: the pre-trained foundational model, named M0, and the five models fine-tuned using the different optimizers presented in Table I, named M1 through M5. Each model was used to generate two sets of 12 images each: one set representing normal cases, created with the prompt “*healthy or normal human chest x-ray*”; and the other representing abnormal cases, created with the prompt “*Human chest x-ray with tuberculosis. Bilateral miliary nodules with Right Middle Lobe infiltrate. Right pleural effusion*”.

Figure 2 and Figure 3 present the sets of images generated for the normal and abnormal cases, respectively. All fine-tuned models (i.e., M1 through M5) were trained for 100 epochs. We presented the generated images to a medical doctor, who evaluated their level of realism using a five-level Likert scale from Very Unrealistic (1) to Very Realistic (5). Normal and abnormal images were evaluated separately for each model.

Table II presents the evaluation of the images made by the medical doctor regarding each model, with respect to their realism. Models M0, M3, M4, and M5 were evaluated as Very Unrealistic (1) for both image sets. Indeed, it can be empirically observed in Figures 2 and 3 that these models generate images that are quite different from a real x-ray image (Figure 1). Additionally, models M4 and M5, in particular, presented a similar result to the foundational model M0 by generating more colorful and cartoon images.

Models M1 and M2 achieved better results, with M1 obtaining a Very Realistic (5) evaluation for normal cases and Average Realism (3) for abnormal cases. Both models

<sup>5</sup>Available at: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

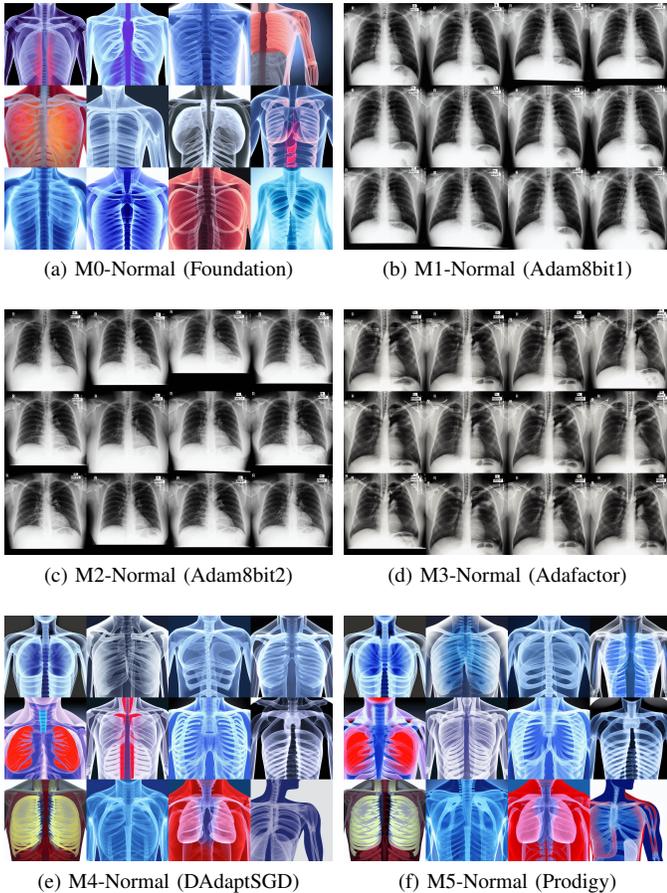


Fig. 2. Set of normal chest x-ray images generated by the models presented in Table I. All models except M1 were fine-tuned for 100 epochs on a dataset of 30 chest x-ray images. All images were generated using the prompt "healthy or normal human chest x-ray".

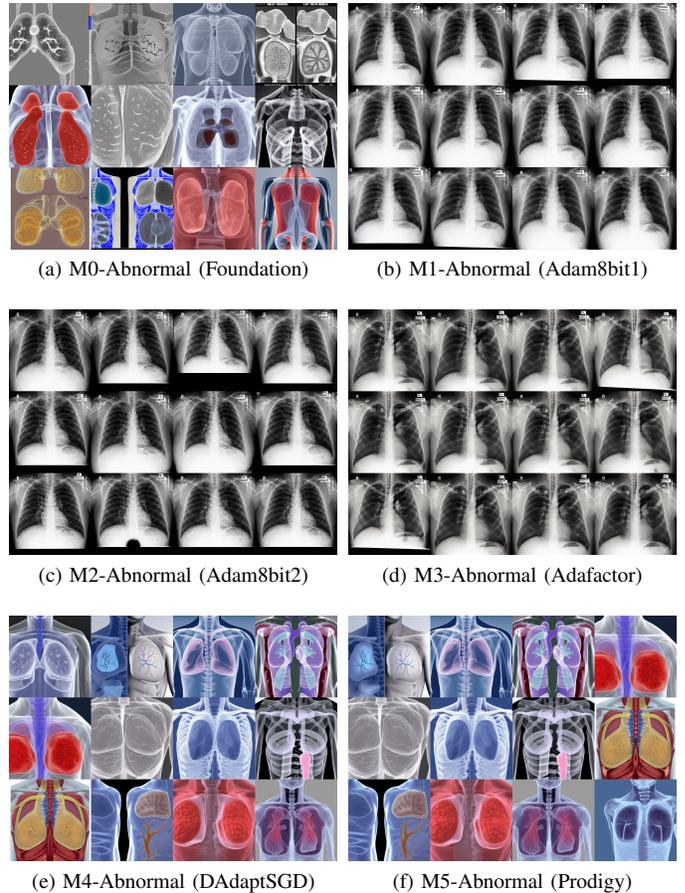


Fig. 3. Set of abnormal chest x-ray images generated by the models presented in Table I. All models except M0 were fine-tuned for 100 epochs on a dataset of 30 chest x-ray images. All images were generated using the prompt "Human chest x-ray with tuberculosis. Bilateral miliary nodules with Right Middle Lobe infiltrate. Right pleural effusion".

utilize the Adam8bit optimizer, with M1 using the default configuration. These results indicate that further experimentation could be conducted using the Adam8bit optimizer, as it demonstrated satisfactory performance. However, it is important to acknowledge the limitation of having only one evaluator.

TABLE II  
RESULTS OF THE MEDICAL DOCTOR'S EVALUATION. IMAGES GENERATED BY EACH MODEL WERE EVALUATED USING A FIVE-LEVEL LIKERT SCALE, FROM VERY UNREALISTIC (1) TO VERY REALISTIC (5).

Model	Normal (Figure 2)	Abnormal (Figure 3)	Total
M0 (Foundation)	1	1	2
M1 (Adam8bit1)	5	3	8
M2 (Adam8bit2)	3	2	5
M3 (Adafactor)	1	1	2
M4 (DAdaptSGD)	1	1	2
M5 (Prodigy)	1	1	2

## V. CONCLUSION

In this work, we presented an initial exploration of the impact of fine-tuning foundation models for generating medical images, focusing on chest X-rays. Our approach considers a Latent Diffusion Model as a base model and different optimizer configurations for the fine-tuning process. We generated a set of images using these models, considering both normal and abnormal cases. Our experiments indicate that even the use of a small dataset for fine-tuning could generate images with satisfactory levels of realism. However, our work has some limitations. The experiments were conducted with input from only one medical doctor and relied on visual inspection. Additional validation techniques could be applied, such as using qualitative metrics to assess the generated images and evaluating potential overfitting in the fine-tuned models.

For future work, we consider further experimentation with different dataset sizes, evaluating the impact of different training times, and evaluating our model with more medical professionals. The Adam8bit optimizer, in particular, presented better results and could be used in such experiments. We also

consider experimenting with additional prompts for abnormal cases, given the variety of conditions that could be presented in an x-ray. Finally, we also consider the development of applications that will be beneficial in both healthcare and educational settings. One concept involves creating an application that allows teachers to use our method to generate personalized examples tailored to their students' needs. This approach can improve the teaching and learning experience by providing a more interactive and engaging environment.

#### ACKNOWLEDGMENTS

This study was partly funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, FAPERGS (RITE CIARS 22/2551-0000390-7), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq). This paper was supported by the Ministry of Science, Technology, and Innovations, with resources from Law No. 8.248, dated October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex, and published in the *Residência em TIC 02 - Aditivo*, Official Gazette 01245.012095/2020-56. The authors would like to thank the medical doctor who participated in our experiments.

#### REFERENCES

- [1] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.
- [2] A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, and A. P. Sales, "Generation and evaluation of synthetic patient data," *BMC medical research methodology*, vol. 20, no. 1, pp. 1–40, 2020.
- [3] C. Thapa and S. Camtepe, "Precision health data: Requirements, challenges and existing techniques for data security and privacy," *Computers in biology and medicine*, vol. 129, p. 104130, 2021.
- [4] I. Keshta and A. Odeh, "Security and privacy of electronic health records: Concerns and challenges," *Egyptian Informatics Journal*, vol. 22, no. 2, pp. 177–183, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1110866520301365>
- [5] N. S. Almaghrabi and B. A. Bugis, "Patient confidentiality of electronic health records: A recent review of the Saudi literature," *Dr. Sulaiman Al Habib Medical Journal*, vol. 4, no. 3, pp. 126–135, 2022.
- [6] S. Sundaram and N. Hulkund, "Gan-based data augmentation for chest x-ray classification," 2021.
- [7] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International conference on machine learning*. PMLR, 2016, pp. 1060–1069.
- [8] H. Dou, C. Chen, X. Hu, Z. Xuan, Z. Hu, and S. Peng, "Pcargan: Incremental orthogonal projection discrimination for face super-resolution," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1891–1899.
- [9] B. Azad, R. Azad, S. Eskandari, A. Bozorgpour, A. Kazerouni, I. Rezik, and D. Merhof, "Foundational models in medical imaging: A comprehensive survey and future vision," 2023.
- [10] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations. naacl-hlt," *arXiv*, vol. Nothing, no. Nothing, 2018.
- [11] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [12] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [14] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8821–8831.
- [15] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.
- [16] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, vol. Nothing, 2021.
- [17] F. Petroni, T. Rocktäschel, P. Lewis, A. Bakhtin, Y. Wu, A. H. Miller, and S. Riedel, "Language models as knowledge bases?" *arXiv preprint arXiv:1909.01066*, vol. Nothing, 2019.
- [18] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "Retrieval augmented language model pre-training," in *International conference on machine learning*, vol. Nothing. PMLR, 2020, pp. 3929–3938.
- [19] Z. Wang, Z. Wu, D. Agarwal, and J. Sun, "Medclip: Contrastive learning from unpaired medical images and text," 2022.
- [20] S. Motamed, P. Rogalla, and F. Khalvati, "Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images," *Informatics in Medicine Unlocked*, vol. 27, p. 100779, 2021.
- [21] W. H. Pinaya, P.-D. Tudosiu, J. Dafflon, P. F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin, and M. J. Cardoso, "Brain imaging generation with latent diffusion models," in *MICCAI Workshop on Deep Generative Models*. Springer, 2022, pp. 117–126.
- [22] W. Chen, H. Hu, C. Saharia, and W. W. Cohen, "Re-imagen: Retrieval-augmented text-to-image generator," *arXiv preprint arXiv:2209.14491*, 2022.
- [23] Y. Zhou, R. Zhang, C. Chen, C. Li, C. Tensmeyer, T. Yu, J. Gu, J. Xu, and T. Sun, "Towards language-free training for text-to-image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 907–17 917.
- [24] H. Ali, S. Murad, and Z. Shah, "Spot the fake lungs: Generating synthetic medical images using neural diffusion models," in *Irish Conference on Artificial Intelligence and Cognitive Science*. Springer, 2022, pp. 32–39.
- [25] K. Packhäuser, L. Folle, F. Thamm, and A. Maier, "Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.
- [26] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *arXiv preprint arXiv:1711.04340*, 2017.
- [27] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.
- [28] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 500–22 510.
- [29] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.
- [30] T. Dettmers, M. Lewis, S. Shleifer, and L. Zettlemoyer, "8-bit optimizers via block-wise quantization," *arXiv preprint arXiv:2110.02861*, 2021.
- [31] H. Face, "8-bit optimizers," 2023. [Online]. Available: <https://huggingface.co/docs/bitsandbytes/main/en/optimizers>
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
- [33] N. Shazeer and M. Stern, "Adafactor: Adaptive learning rates with sublinear memory cost," 2018.
- [34] A. Defazio and K. Mishchenko, "Learning-rate-free learning by d-adaptation," in *International Conference on Machine Learning*. PMLR, 2023, pp. 7449–7479.
- [35] K. Mishchenko and A. Defazio, "Prodigy: An expeditiously adaptive parameter-free learner," *arXiv preprint arXiv:2306.06101*, 2023.