

Generating and Comparing AI-generated Biblical Images

Hide Makimei¹[0009-0004-7184-7943], Shuai Wang¹[0000-0002-1261-9930], and
Willem van Peursen³[0000-0002-3142-5752]

¹ Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

`hide.n.g.makimei@student.vu.nl | shuai.wang@vu.nl`

² Eep Talstra Centre of Bible and Computer (ETCBC), Faculty of Religion and Theology, Texts and Traditions, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

`w.t.van.peursen@vu.nl`

Keywords: Biblical images · generative AI · Dall E · stable diffusion · religion.

1 Introduction

Despite the fact that more and more Artificial Intelligence (AI) tools can generate art, the generated images have been argued to lack human attributes such as creativity, originality, subjectivity, emotional depth, context, cultural significance, intention, and conceptualisation [2, 3, 5]. A comparison of AI-generated images with well-known human paintings of the same topic can help understand the confounding differences between AI and humans, whilst analyzing the bias of generators and guiding future AI-based tools. This comparison could also help the selection of relevant images and ease manual evaluation. The biblical text has served as a wellspring of inspiration for human creativity across various domains. Its stories, metaphors, ethical teachings, and representations of divine creation have guided and fueled the imagination of artists. With artistic expression, moral reflection, or innovative problem-solving, the connection between the biblical text and human creativity remains a profound and enduring one, making it an interesting perspective of how AI would illustrate the same prompt. Given that humans are often present in biblical stories and we can take advantage of existing tools for human detection and facial recognition, we focus on the human beings in the generated images and compare how they vary between generators and how they differ from masterpieces in art history. Since the evaluation of generated images at a large scale is not feasible for manual examination, we employ some Machine Learning models for image assessment. This thesis focuses on generating and evaluating AI-generated images using biblical text as a reference and comparing them to human art. Several features were evaluated, including aesthetics, accuracy regarding religious background, and the human characters created.

2 Data and Methodology

We provide a new dataset that consists of 7,116 images from 3 generators Midjourney, Dall E 2, and seven variants of Stable Diffusion. We use 5 passages from the Bible (the King James version) as prompts. The results are compared to human artworks from the Renaissance for similar ideology, painting style, and biblical understanding. Both the generated data and selected paintings are passed through CNNs that classify the characters' age, gender, and the whole image's sentimental value. The process then transforms the classification into interval values, which makes us able to compare the values numerically. The age-gender classification uses two pretrained models Detectron2 [6], a Mask R-CNN with a ResNet50 + FPN as the backbone. Age model classifier, which is a LeNet-5, consists of 3 convolutional layers and 2 fully connected layers developed by Gil Levi and Tal Hassner [4]. Sentimental classification is an AlexNet-styled network comprising five convolutional layers and three fully-connected layers [1]. The score was calculated by taking the numeric value of the targeted human artwork produced by the given CNN model (e.g. number of people detected in interval form). As a proof-of-concept, the values are combined using a simple method that takes the average of all aspects assessed for all the generated images. The value is then compared with its AI counterpart. The results are scored in an interval format of 0 to 1. The lower the overall score, the more similar the images are compared to human art.



Fig. 1. Selected images about the last supper by Midjourney

3 Evaluation

Finally, we compare the evaluation results with the above-mentioned metrics. The overall score for Midjourney is 0.127, and that of Dall E is 0.017. The scores for variants of Stable Diffusion vary between 0.144 and 0.156. Thus, the images by Midjourney are the most similar to human art. Figure 1 are such examples by Midjourney. The AI images are also analyzed for their aesthetic and biblical aspects such as accuracy and style to understand the AI's patterns in the field of theology and art. Furthermore, this thesis attempts to explore the use of image-generating AI in the field of theology. It proposes further avenues in studying different scriptures (translations, versions, etc.) in AI image generation to explore this cross-disciplinary study.

References

1. Victor Campos, Brendan Jou, and Xavier Giró-i-Nieto. From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction. *CoRR*, abs/1604.03489, 2016.
2. Anjan Chatterjee. Art in an age of artificial intelligence. *Frontiers in Psychology*, 13:1024449, 2022.
3. Mingyong Cheng. The creativity of artificial intelligence in art. In *Proceedings*, volume 81, page 110. MDPI, 2022.
4. Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*, June 2015.
5. Bai Liu. Arguments for the rise of artificial intelligence art: Does ai art have creativity, motivation, self-awareness and emotion? *Arte*, Avance en línea:1–11, 04 2023.
6. Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. 2019.