SCHOLARLY REVIEW

Comparing AI and humans' ability to recognize AI-generated images

By Yutong Zhang

Author Biography

Yutong (Catherine) Zhang is a sophomore at Sage Hill School in Newport Coast, California. Catherine aspires to enter the STEM field in the future, and she is passionate about technology education and ethics in AI.

Abstract

As AI technology develops rapidly, AI-generated realistic images have become increasingly deceiving to the human eye. This study compares humans' ability to recognize AI-generated images to that of AI models designed to detect these images. The study prompted humans and machines to classify 80 images of four categories (animal, architecture, landscape, and vehicle) as real or generated. Experiment results show that both machines and humans performed better in recognizing real images than generated images. While humans and machines performed similarly in correctly identifying generated images, machines have significantly higher accuracy when identifying real images. However, the variation among individual machines' accuracies is greater than that among humans. Humans' perception of reality was a key factor that influenced their performance both positively and negatively, as it helped people discern unrealistic traits but also tricks the human eye. Additionally, certain limitations of text-to-image generation models, such as failure to generate letters and eye reflections, aided humans' performance but not machines' performance.

Keywords: AI-generated images, Generative AI, Text-to-image models, Image authenticity, Ethics in AI, Artificial intelligence, Computer science, Human perception, Survey

Introduction

AI image generation has become increasingly popular due to the fast development of AI technology and chatbots. By December 2023, an estimated “250 million companies (77%) are using or exploring AI in some capacity” (Roller, 2023). Some popular AI-image generation models include Dall-E, Stable Diffusion, and Midjourney (MIT Sloan Teaching & Learning Technologies, n.d.). Following the release of these models, potential problems have emerged. The AI models may infringe artists’ copyrights and unintentionally plagiarize their works since AI models train on existing artworks (Walsh, 2023). As generative AI technology advances, these AI-generated images are also becoming increasingly deceiving to the human eye.

The present paper compares the performance of AI-generated image detectors with that of humans when classifying real and AI-generated realistic images. Overall, machines outperformed humans on this task. While machines' performance in recognizing generated images is similar to that of humans, they demonstrated significantly higher accuracy in identifying real images. Both machines and humans achieved higher accuracy in correctly identifying real images than generated images.

Literature review

In a previous study by Nighting & Farid (2022), the researchers asked 315 participants to classify 128 photos of human faces as real or AI-synthesized, and the participants reached an accuracy of 48.2%. Their research focused on humans’ performance in discerning generative AI and demonstrated how AI-generated images trick humans’ perceptions.

Other previous studies focused on training detection models to classify AI-generated images. Baraheem and Nguyen (2023) trained multiple detection models to detect GAN images, and their best model reached 100% accuracy when tested with a 12,000 images dataset. However, GAN image detectors may not work with the same accuracy when detecting generated images from models from different architectures. As shown by a study by Epstein et al. (2023), detectors trained with images from models of one architecture are less effective when classifying images from models of different architectures.

Additionally, Lu et al. (2024) tested both detection models trained from large datasets and humans to classify real and generated images. They found that humans performed worse than detectors overall, getting 38.7% of the questions wrong while detectors only got 13.7% wrong.

Methods

This study used 80 images, consisting of 40 authentic and 40 generated images. The images were divided into four categories — animal, architecture, landscape, and vehicle. Real images were gathered through news sources such as CNN News and BBC News to increase the likelihood of their authenticity. AI-generated images were created with Google Gemini, a newer model released on Dec 06, 2023 (Pichai, 2023), to challenge AI-generated image detectors designed for popular text-to-image models.

In a survey, fifty people classified the 80 photos, arranged in randomized order, as real or generated. Survey participants were gathered through Prolific — an online platform for finding research participants. The participants did not need prior qualifications, and personal information was not collected during the research process. Additionally, this study tested four generated image detection tools — Illuminarty, SightEngine, Is it AI, and Hive Moderation. These tools were selected from internet searches; tools that are free of charge were considered with priority due to a limited budget.

Results

Survey

Figure 1

Survey Scores Distribution

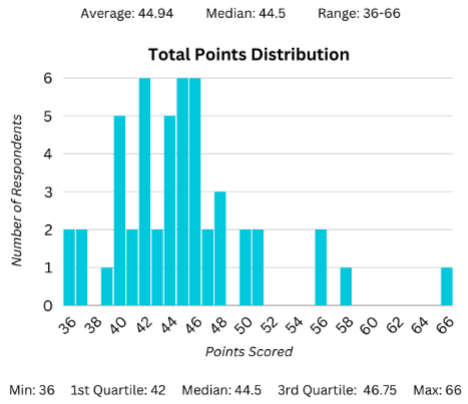

Table 1
Humans' Accuracy by Categories

Category	Real	Generated	Average
Animal	70.2%	36.0%	53.1%
Architecture	70.6%	48.8%	59.7%
Landscape	64.0%	42.2%	53.1%
Vehicle	68.8%	48.8%	58.8%
Average	68.4%	44.0%	56.2%

The survey participants' mean score is 44.9 (56.2%), and the median score is 44.5, showing a clustered distribution with a few outliers. The survey scores are centered around the median score of 44.5, with an interquartile range of 4.75 points and a standard deviation of 5.65 points (7.06%) (Figure 1).

Humans achieved an accuracy of 56.2% overall, with a higher accuracy of 68.4% when identifying real images than an accuracy of 44.0% for generated images. All categories except for the animal category resulted in an accuracy difference of less than 25% between their real and generated images. Humans' accuracy for real animal images is 70.2%, one of the higher accuracies of all categories for real images. However, their accuracy for AI-generated animal images is the lowest of all categories, only reaching 36.0% (Table 1). The range of the overall accuracies for each category is 6.6%, with the architecture category having the highest accuracy of 59.7%, followed by vehicle at 58.8%, animal at 53.1%, and landscape at 53.1%.

AI-generated image detectors

Table 2
AI Tools' Accuracy by Categories

Category	Real	Generated	Average
Animal	97.5%	30.0%	63.8%
Architecture	97.5%	67.5%	82.5%
Landscape	97.5%	42.5%	70.0%
Vehicle	95.0%	42.5%	68.8%
Average	96.9%	45.6%	71.3%

Table 3
Humans' Accuracy vs. AI Tools' Accuracy

Subject	Real	Generated	Overall
Human	68.4%	44.0%	56.2%
AI Tools	96.9%	45.6%	71.3%

AI tools reached an overall accuracy of 71.3%. The machines have a notably higher performance in recognizing real images than humans, reaching an average accuracy of 96.9% in real images compared to that of 68.4% for humans. However, AI tools performed similarly as humans when detecting generated images — the former achieved a 45.6% accuracy, while the latter achieved 44.0% (Table 3). Distribution-wise, the accuracies for the four AI tools tested — Illuminarty, SightEngine, Is it AI, and Hive Moderation — are respectively 75.0%, 63.8%, 60.0%, and 86.3%, with a standard deviation of 11.9%. By category, machines have the highest accuracy of 82.5% in architecture, followed by 70.0% in landscape, 68.8% in vehicles, and the lowest accuracy of 63.8% in the animal category (Table 2).

Discussions

Variation in machines' performance

Table 4
Accuracy by AI Tools

Model	Real	Generated	Average
Illuminarty	95.0%	55.0%	75.0%
SightEngine	100%	27.5%	63.8%
Is it AI	92.5%	27.5%	60.0%
Hive Moderation	100%	72.5%	86.3%

AI tools demonstrated a greater variation in performance than humans — the former has a standard variation of 11.9%, while that of the latter is 7.06%. Specifically, AI tools' performance vary significantly for generated images, ranging from accuracies of 72.5% to 27.5% (Table 4). However, the reason behind the variation of their ability remains unclear. SightEngine (n.d.) and Hive Moderation (n.d.) both claim that their models detect generated images from a number of popular image generators listed on their documentation, in which Gemini is not included. However, the former has an accuracy of 27.5%, but the latter performed relatively well with images generated by Gemini, reaching an accuracy of 72.5%. While Illuminarty (n.d.) and Is it AI (n.d.) don't state that their models are limited to certain AI image generators, they did not perform as well as Hive Moderation, having accuracies of 55.0% and 27.5%, respectively.

Humans' perception of reality

Humans' perception of reality can both aid and hinder their ability to recognize AI-generated images. In some cases, humans misclassify real and generated images due to humans' tendency to perceive certain scenes/situations as real or fake. In other cases, however, this tendency can help humans identify unrealistic traits of AI-generated images.

Figure 2
Real Animal Photo used in the Survey

Human Accuracy: 5/50 Machine Accuracy: 3/4

Humans have a higher chance of incorrectly categorizing real photos as generated when the photo is of a moment or perspective that is hard to capture through the human eye, making the photo seem unrealistic. This condition is satisfied in many cases for wildlife photography. For example, Figure 2 is an authentic photo taken by wildlife photographer Thomas Vijayan. In the survey, 5 out of 50 people correctly identified the photo as real. According to Thomas, he "kept the camera settings fast enough to freeze the running tiger" (Cornall, 2023), capturing a moment that the human eye cannot capture. Therefore, humans perceived this image as unrealistic due to their limited perspectives of reality.

Figure 3
Generated Landscape Photo used in the Survey

Human Accuracy: 10/50 Machine Accuracy: 3/4

On the other hand, humans may also perceive generated images as realistic when an image resembles properties of the natural world. One example is Figure 3, a generated landscape image that only 10 out of 50 participants correctly classified. In this image, the river appears narrower as it travels further, employing the artistic principle of perspective, which recreates the three-dimensional property of nature (Tate, n.d.). Since the image resembles a major characteristic of the natural world, humans are more likely to perceive it as realistic.

Figure 4
Left - Generated Architecture Image, Right - Real Architecture Image

Human Accuracy: 32/50
Machine Accuracy: 4/4

Human Accuracy: 34/50
Machine Accuracy: 4/4

Humans reached an accuracy of 48.8% in generated architecture images, one of the highest of all categories for generated images. Generated architecture images often lack background details, while real images show the surroundings of the architecture, as demonstrated by the pair of images shown in Figure 4. Studies by psychologists suggest that human perception is “strongly influenced by context” (Seth, 2022). Therefore, authentic images are perceived as more realistic by the human eye since their prolific background details establish the environmental context of the architecture, increasing the plausibility of its existence. In this case, humans’ perception of reality, influenced by the images’ context, helps humans correctly identify real and generated images.

Limitations of text-to-image models

Figure 5
Generated Vehicle Image used in Survey

Human Accuracy: 25/50 Machine Accuracy: 1/4

One major limitation of AI-image generation models is that they perform poorly in generating letters, often resulting in spelling errors and misshaped letters (Growcoot, 2024). Humans can quickly identify unclear letters or wrong spelling through common sense. However, machines cannot easily identify spelling errors since AI image detectors, trained with image data, search for patterns in images rather than patterns in text. This differentiation is reflected as humans outperformed machines in identifying AI-generated images of vehicles — humans reached an accuracy of 48.8%, while machines had an accuracy of 42.5% (Table 1 & Table 2). Many generated vehicle images contain letters with unclear spelling on the vehicles' license plates. For example, in Figure 5, the letters on the top plate of the taxi are barely spelled out. As a result, 25 out of 50 people correctly identified Figure 5 as generated, while only 1 out of 4 machines responded correctly.

Figure 6
Generated Animal Image used in the Survey

Human Accuracy: 16/50 Machine Accuracy: 0/4

Humans also performed slightly better than machines in the generated animal images category, achieving accuracies of 36.0% and 30.0%, respectively (Table 1 & Table 2). A recent discovery by Tonkin et al. at the Royal Astronomical Society (2024) reveals that AI image generators have trouble generating reflections in the eyes — the reflections in a pair of eyes are often mismatched (Tonkin, 2024). This is the case for Figure 6, where the animal's left eye has a reflection on the left side, but the right eye has almost no reflection. In addition, both eyes are pure black with no combination of other colors, which is typically unrealistic. Sixteen out of 50 people correctly classified this image as generated, while 0 out of 4 AI detection tools did so. Their performances reflect that humans more efficiently observe unrealistic eye reflection patterns, but machines will need more training to achieve this.

The limitations of text-to-image models discussed above aided humans' performance since they result in image details that are contrary to common sense. However, machines' performance didn't benefit from these limitations, showing that machines either have difficulties or lack training in recognizing these patterns.

Conclusion

Machines performed better overall than humans in classifying images as real or generated. Specifically, humans and machines performed similarly in correctly identifying generated images, but machines are notably better at accurately identifying real images. Multiple factors, such as humans' perception of reality and the limitations of text-to-image models, influenced the study results. However, this study does not represent the performance of all humans and AI detection tools; it aims to provide insights into humans and AI's ability to recognize generated images. Generated images can easily deceive humans and machines, as shown by their accuracies of less than 50% in the generated images category. It is significant for humans to be aware of generative AI's ability to fool both the human eye and AI algorithms so that false information does not spread through AI-generated images.

References

Baraheem, S. S., & Nguyen, T. V. (2023). AI vs. AI: Can AI Detect AI-Generated Images?. Journal of Imaging, 9(10), 199.

Cornall, F. (2023, April 4). "The New Big 5": Wildlife photography book turns the lens on animals at risk. CNN. https://www.cnn.com/travel/article/new-big-5-wildlife-photography-book-c2e-scn-spc-intl/index.html

Epstein, D. C., Jain, I., Wang, O., & Zhang, R. (2023). Online detection of ai-generated images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 382-392).

Growcoot, M. (2024, March 6). Why AI image generators struggle to get text right. PetaPixel. https://petapixel.com/2024/03/06/why-ai-image-generators-struggle-to-get-text-right/

Hive Moderation. (n.d.). Machine learning models to detect AI-generated content. https://hivemoderation.com/ai-generated-content-detection

Illuminarty. (n.d.). Is an AI behind your image? https://illuminarty.ai/en/

Is it AI. (n.d.). How it Works. Is It AI? https://isitai.com/how-it-works/

Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X., & Ouyang, W. (2024). Seeing is not always believing: benchmarking human and model perception of AI-generated images. Advances in Neural Information Processing Systems, 36.

MIT Sloan Teaching & Learning Technologies. (n.d.). AI Image Generation Tools. https://mitsloanedtech.mit.edu/ai/tools/images/

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8). https://doi.org/10.1073/pnas.2120481119

Pichai, S. (2023, December 6). Introducing Gemini: Our largest and most capable AI model. Google. https://blog.google/technology/ai/google-gemini-ai/

Roller, J. (2023, December 11). Ethical pros and cons of AI Image Generation. IEEE Computer Society. https://www.computer.org/publications/tech-news/community-voices/ethics-of-ai-image-generation

Seth, A. (2022, October 3). The big idea: Do we all experience the world in the same way?. The Guardian. https://www.theguardian.com/books/2022/oct/03/the-big-idea-do-we-all-experience-the-world-in-the-same-way

Sightengine. (n.d.). AI-Generated Image Detection. https://sightengine.com/docs/ai-generated-image-detection

Tate. (n.d.). Perspective coursework guide. https://www.tate.org.uk/art/student-resource/exam-help/perspective

Tonkin et al. (2024, July 17). Want to spot a deepfake? Look for the stars in their eyes. The Royal Astronomical Society. https://ras.ac.uk/news-and-press/news/want-spot-deepfake-look-stars-their-eyes

Walsh, D. (2023, August 28). The legal issues presented by Generative AI. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai