Measuring Generative Artificial Intelligence Understanding of Human Psychology and What Makes People Scared

Sean T. Slaughter
Clemson University
Clemson, SC, USA
stslaug@gmail.com
stslaug@clemson.edu

(a) Image generated with Google's Imagen 3 model. Given guidance to generate something generically cute

(b) Image generated with openAI's Dall-E model. Given guidance to generate something generically scary

Figure 1: Examples of generated images

ABSTRACT

This study explores the human gaze and how fear-inducing and calming stimuli can provoke instinctual reactions. With the increasing availability of generative artificial intelligence (Gen AI), there are rising concerns [Varnum 2023]. In this study, we examine what these creations of people, Gen AI, can be used to elicit emotional responses. Utilizing four free Gen AI models, we generated images and presented them to participants. Our objective was to test if the AI-generated images provoked a sense of unease and anxiety and if there was a calming or no change from intentionally "cute" prompted images. Using eye-tracking data is used to analyze the gaze patterns of the participants and evaluate the emotional impact of stimuli.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**.

ETRA '23, May 29–3 June, 2023, Tübingen, Germany

ACM ISBN 978-1-4503-8344-8/21/05
https://doi.org/10.1145/3448017.xxxxxxx

KEYWORDS

eye tracking, visual attention, cognitive load

ACM Reference Format:
Sean T. Slaughter. 2024. Measuring Generative Artificial Intelligence Understanding of Human Psychology and What Makes People Scared. In *2023 Symposium on Eye Tracking Research & Applications, May 29–3 June, 2023, April 23–28, 2023, Tübingen, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3448017.xxxxxxx

1 INTRODUCTION

Technology is rapidly expanding worldwide, especially with the emergence of publicly accessible generative artificial intelligence (Gen AI) models. While these Gen AI models, such as Dall-E and Meta AI, have gained popularity due to their usefulness and ease of use. They have also had a rising fear of whether they are safe or ethical. [Cross 2023]

With this study, we are pushing those fears directly, exploring if Gen AI can generate anxiety or fear-inducing images. We investigated if these Gen AI models can accurately and consistently predict what can elicit these negative emotional responses. To test this, eye-tracking technologies track even minute changes within a participant. With eye tracking, we can track subtle responses and see the visual and emotional engagement of the participants, as well as even signs of anxiety or unease.

In this work, we generated a diverse set of images and presented

them to volunteer participants to track and record their eye movements. By analyzing this data, we will assess the degree of anxiety or fear induced by the individual participant. This approach will test both the effectiveness of eye-tracking in identifying human fear and anxious responses and the Gen AI model's ability to understand emotional reactions.

This data can be used to understand Gen AI and its limitations better. It also points out how effectively eye tracking can track or indicate emotion.

2 BACKGROUND

Using generative methods to create visually stunning art isn't a new topic, especially with today's emergence of easily accessible image generation. However, a Media Lab team at Massachusetts Institute of Technology (MIT) used machine learning to provoke fear. The project was named the Nightmare Machine. The Nightmare Machine explores the question of human fear and whether deep learning and computers understood what made humans have anxiety and other adverse effects [Yanardag et al. 2021].

This study received hundreds of thousands of participants on testing across 147 countries, resulting in over one million evaluations. This gave the researchers tremendous data to test and prove their generation's effectiveness. Overall, they found that increasing exposure to images significantly increases adverse effects and states of anxiety. This study plans to utilize the same ideas. However, it will be validated with eye-tracking to validate the emotional or anxious response of participants.

3 EMPIRICAL VALIDATION

We conducted experiments to evaluate humans' and their eye's reactions to generative artificial intelligence (GenAI). Specifically, GenAI can generate a variety of cute and scary images.. . .

3.1 Experimental Design

This study tests how the human eye reacts to Generative AI's ability to predict human fear and what participants do in the face of unsettling and calm images. Using eye-tracking tools and software, in this study, we document and process minute movements and reactions in a person, focusing on gaze patterns. Focusing on their fixations, saccades, and rapid eye movements, we can verify if an individual feels a desired anxiety or response to fear. This will allow my testing to gather data and unknown reactions from participants.

The variables influencing the experiment are whether the image was generated for its creepy or cute nature and the generation model. These two variables will be taken into account when viewing a participant's fixations, area of interest, attention, saccades, and response. We hypothesize that participants will exhibit more pronounced physiological responses to creepy images and will show signs of anxiety over cute photos. Furthermore, We predict that participants will focus more on pleasant images while avoiding looking directly at unsettling images.

To mitigate context biases within participants, all images will be randomly presented in different orders from other participants. This is to ensure that previous stimuli don't influence the reaction of stimuli.

3.2 Participants

[Not Sure Yet / Not needed for proposal]

3.3 Procedure

This section will review how data was collected and our image generation methodology within this study.

3.3.1 *Image Generation and Selection*.

This study experiment utilizes four free and popular generative artificial intelligence (Gen AI) image models. The language models used are Microsoft Co-pilot's DALL-E 3, OpenAI's DALL-E 3, Google's Imagen 3, and Meta's Emu model. We will generate eight images to represent cute and creepy images according to AI to encompass and show off each model thoroughly. This spread of images leaves three images for each category and each model.

Potentially cute and pleasant images could include pictures of animals (e.g., puppies, kittens), aesthetic scenes (e.g., Flowers or sunrises), or people. Negative or potentially unsettling images may contain unfamiliar creatures, dark environments, or popular phobia depictions. Each is intended to hit at least one image per category that correlates with the intended feeling. For each type, cute or creepy, We use the same prompt for each language model. After generating images based on each prompt, we chose the first image provided if multiple images were generated.

Aside from generation categories, we ensure the format quality of the images is consistent. This includes ensuring that each image is cropped to the exact resolution as the others. We are cropping over scaling our stimuli to avoid deformations that may arise or influence a participant's experience.

3.3.2 *Data Collection*.

In this work, we focus on eye tracking and how the eye reacts to different stimuli. To validate a response in the individual, I will be testing and collecting how focused an individual is on an area, any fixations in other regions, self-reporting, and rapid eye movements.

Areas of interest (AOI)

In this study, I will be utilizing AOIs to mark and check if specific regions are being noticed more often by participants. Specific regions may be objects, dark areas, and high contrast levels. AOI locations will be chosen based on visual prominence (e.g., contrasting colors and edges). The number of AOIs will also change for each image, while the rules I apply stay the same. Depending on the complexity and number of objects, I may use less or more, depending on the degree of each. When choosing, I will focus on faces, the center of the image, threatening elements (e.g., weapons, predatory animals, teeth), violent scenes, unexpected elements (e.g., a cat in a cemetery), and areas with high contrasts. These are the top priority for AOI placement, but more will be placed to fill in other missing cases that may not fit those categories.

I also will keep track of other fixation points in the stimuli. If they don't adhere to a defined AOI, participants' fixations will be compared with one another to find the correlation. **Saccades**

When people make rapid, involuntary eye movements, it causes a slight and immediate shift from one point to another. There are four main measurements for saccades: amplitude, duration, velocity, and latency. Amplitude, about saccades, is the angular distance between the starting and ending points. The duration is the time it

Figure 2: Bing's Dall-E 3 Model's generated creepy image

takes to complete the saccade. Velocity is the speed of the saccade over its distance. Finally, latency is the time between exposure to stilumi and the saccade initiation.

Self-Reported Ratings

To collect how the participant reacted emotionally and internally, I will ask each participant how an image may have made them feel after each stimulus. I will use a one to ten-point rating system ranging from cute, with no issues with looking at it, to scared, or can't continue to look at it. The numbers will be removed from the factor so the user isn't taunted about the option or picking a perfect number. To implement this scale, we will provide a slider for the user. To ensure the input information is accurate and uninfluenced, I will also not be watching the participants directly to ensure they aren't worried about external judgment.

3.4 Apparatus

A Gazepoint GP3 eye tracker was used, sampling at 60hz with an accuracy of 1°visual angle as given by the manufacturer, see Figure 2

4 RESULTS

This section presents the findings from our experiment examining participants' eye-tracking responses and self-reported emotional reactions to generative AI-produced stimuli categorized as "cute" or "creepy." The primary objective was to determine how individuals can react to stimuli that is intended to be calming or scary. We analyzed key metrics to assess these outcomes, including fixation counts, fixation durations, emotional ratings, reaction times, and survey responses. A total of nine participants were involved in the study.

4.1 Fixation Patterns

Number of Fixations. When viewing cute images, participants exhibited more distributed fixation patterns across multiple areas of interest (AOIs). The mean number of fixations per participant on cute images was approximately 13 (SD = 2). The disparity in unique AOI fixations averaged three, indicating a holistic processing approach. For example, in the image generated by OpenAI's DALL-E 3 language model intended to be cute (hereafter referred to as the "OpenAI cute image"), fixations were spread evenly across AOIs. This suggests that 80% (7 out of 9) of participants engaged with the overall composition rather than focusing on specific elements.

In contrast, fixations were concentrated on specific AOIs in creepy images, with a higher disparity in fixation counts. The mean number of fixations per participant on creepy images was approximately nine (SD = 3). For instance, the intended creepy image generated by Bing's DALL-E model (the "Bing creepy image") recorded an average of nine fixations on one AOI and three on another, resulting in a disparity of six. This concentration shows heightened emotional engagement, which may be driven by anxiety, discomfort, or fear. Notably, 67% (6 out of 9) of participants focused on the most salient AOI in creepy images.

4.2 Fixation Duration

Cute Stimuli. Participants focused less on each AOI in cute images, indicating a more relaxed cognitive state. The mean fixation duration per AOI for cute images was approximately 1.5 seconds (SD = 0.3). For example, the OpenAI cute image showed shorter fixation durations across all AOIs, reflecting minimal emotional arousal and a calm engagement with the content.

Creepy Stimuli. Participants exhibited longer fixation durations on specific AOIs in creepy images. The mean fixation duration per AOI for creepy images was approximately 2.2 seconds (SD = 0.5), which is higher than that of cute images. In the creepy image generated by OpenAI's DALL-E 3 model (the "OpenAI creepy image"), AOI one had consistently higher fixation times, with 80% (7 out of 9) of participants fixating on it for more than two seconds. This attention could suggest that creepy images require more mental processing, likely due to their intentionally unsettling nature.

4.3 Emotional Ratings

Cute Stimuli. Participants consistently rated cute images highly on the cuteness scale. On a scale from one to ten, the mean rating was 9.0 (SD = 0.8), with 87% (8 out of 9) of participants rating them eight or higher. This aligns with the intended calming effect of these stimuli. However, it is noteworthy that the cute images generated by Meta's Emu and Google's Imagen language models contained visual artifacts, such as deformed eyes in the dogs and the human family depicted. This issue caused 53% (5 out of 9) of participants to rate these specific cute images as cute and creepy, with creepiness intensities all above five on a scale of one to ten.

Creepy Stimuli. Creepy images received high scariness ratings, with a mean score of 8.5 (SD = 0.9). On the same one to ten scale, 87% (8 out of 9) of participants rated them seven or higher. These ratings confirm that the generative AI-produced creepy stimuli successfully elicited the intended negative emotional responses. In the surveys, 80% (7 out of 9) of participants reported feelings of anxiety, fear, or discomfort when viewing the creepy images, specifically noting unsettling elements such as unnatural shapes and dark contrasts.

4.4 Reaction Times

Cute Stimuli. Reaction times were generally longer for cute images. The mean reaction time for cute images was approximately 2.5 seconds (SD = 0.5). The OpenAI cute image elicited a mean reaction time of nearly 3 seconds, suggesting a more relaxed processing state consistent with the calming nature of cute stimuli.

Creepy Stimuli. Reaction times were shorter for creepy images, with a mean of approximately 1.8 seconds (SD = 0.4). Bing's creepy image had a mean reaction time of approximately 1.5 seconds. Notably, 67% (6 out of 9) of participants had reaction times under 2 seconds for creepy images.

4.5 Variability in Fixation Patterns

Creepy Stimuli. There was a large range in fixation counts and fixation durations for creepy images. This shows that some participants avoided disturbing areas while others focused more intensely on them. 40% (4 out of 9) of participants fixated on figures and distinct objects for more than two seconds, 33.3% (3 out of 9) viewed all AOIs with near equal timings, while 27% (2 out of 9) avoided figures altogether. Survey responses indicated that while some participants felt compelled to examine unsettling features closely, others preferred to avert their gaze.

Cute Stimuli. Fixation behaviors were more consistent across participants when viewing cute images, indicated by lower standard deviations in fixation counts and durations. This uniformity suggests that the cute images consistently elicited similar individual engagement patterns. Over 87% (8 out of 9) of participants displayed distributed fixation patterns across all AOIs.

4.6 AOI Relevance

Creepy Stimuli. Fixations were consistently toward specific AOIs featuring visually prominent elements such as unnatural shapes, dark contrasts, or threatening features. This targeted attention may be a cause for investigating the effectiveness of the AI's images or the prominence of the figure. Participants reported in surveys that these elements were particularly engaging, albeit in a discomforting manner. Specifically, 73% (7 out of 9) of participants identified these AOIs as the most attention-grabbing aspects of the images.

Cute Stimuli. Participants explored AOIs more evenly in cute images, reflecting an interest in the overall composition and reinforcing the perception of calmness and wholesomeness. However, due to the visual artifacts in the cute images generated by Emu and Imagen, such as deformed eyes, participants reported mixed reactions. Half of the participants (50%, 5 out of 9) rated these images as cute and creepy, indicating that the artifacts introduced elements of discomfort. Over 87% (8 out of 9) of participants appreciated the overall aesthetic of the cute images without artifacts.

4.7 Survey Responses

4.7.1 General Impressions. Participants generally found the images engaging, with 80% (7 out of 9) selecting "Engaging" or "Very Engaging" on the provided scale. The images were perceived as realistic, with 87% (8 out of 9) indicating "Somewhat Realistic" or "Very Realistic." Several participants noted particular elements or artifacts that caught their attention, such as the lifelike textures in cute images or unsettling distortions in creepy images. Approximately 60% (5 out of 9) of participants reported noticing specific details that enhanced their emotional responses.

4.7.2 Emotional Responses. When asked to measure their emotional response to the cute images, 87% (8 out of 9) of participants chose "Positive" or "Very Positive," reinforcing the calming effect observed in the eye-tracking data. However, for the cute images generated by Emu and Imagen, 53% (5 out of 9) of participants reported mixed feelings due to the visual artifacts, rating them as both cute and creepy, with intensities of participants rated above 5.

On the other hand, responses to the creepy images were mostly "Negative" or "Very Negative." With 87% (8 out of 9) of participants agreeing, it points to our hypothesis that creepy images come with increased fixation durations and faster reaction times.

Most participants (80%, 7 out of 9) acknowledged that specific images caused feelings of anxiety, fear, or discomfort, explicitly citing the creepy images generated by Bing's DALL-E and Meta's Emu language models. They described elements such as eerie lighting, distorted faces, and unnatural movements as contributing factors.

4.7.3 AI Model Effectiveness. Participants generally believed that the images were likely generated by artificial intelligence, with 73% (7 out of 9) selecting "Likely AI" or "Most Likely AI." They agreed that the creepy images accurately depicted elements that can induce fear or anxiety, with 87% (8 out of 9) selecting "Agree" or "Strongly Agree." Similarly, 87% (8 out of 9) agreed that the cute images effectively elicited positive emotions, although some expressed reservations about the images with artifacts.

On a scale of one to five, assessing the effectiveness of AI-generated images at conveying emotions, the mean rating was 4.0 (SD = 0.6), indicating participants found them "Effective" to "Very Effective." This feedback supports the conclusion that generative AI models are capable of producing emotionally impactful content, though artifacts can affect perceptions.

4.7.4 Reflections. When asked if anything unexpected occurred during the experiment, some participants (33%, 3 out of 9) mentioned being surprised by the realism and emotional impact of the AI-generated images. Our belief in this is that participants noted that visual artifacts displayed in Meta and Google's cute images affected their emotional responses negatively.

5 DISCUSSION

This study examined whether generative AI models can create images that elicit specific emotional responses, such as calmness or unease, and analyzed participants' eye-tracking data. The findings suggest that Gen AI models can influence engagement, and eye-tracking metrics provide valuable metrics into reactions to almost any digital content.

5.1 Emotional Responses and Eye-Tracking Metrics

Participants' ratings and eye-tracking data indicate that Gen AI models can produce images that evoke targeted emotions. Cute images received high cuteness ratings, longer reaction times, and distributed fixation patterns, reflecting relaxed processing. Creepy

Figure 3: Meta's Emu language model's first generated cute intended image. Depiction used for it's artifacts in the figure's eyes.

Figure 4: Google's Image 3 language model's first generated cute intended image. Depiction used for it's artifacts in the figure's eyes.

images received higher anxiety ratings, shorter reaction times, and concentrated fixations on specific AOIs, indicating heightened emotional intensity.

In some generated images, artifacts (Figure 2) led to mixed emotional responses, with participants rating specific images as cute and creepy rather than one or the other. This points to AI limitations, where generation errors can have an emotional impact. This issue can be improved through more development time for these models and increased training.

5.2 Variability in Visual Engagement

Cute images prompted a large range of fixations across AOIs, aligning with their calming nature. Generated creepy images led to concentrated fixations on unsettling elements. Some participants focused intensely on disturbing features, while others avoided them, indicating individual differences in tolerance for discomfort. This just shows that there was a range of fixation patterns among participants.

5.3 Implications and Future Work

Our study shows that Gen AI models can create emotionally influential images, but effectiveness depends on image quality. The presence of artifacts suggests that current generative AI models could improve. With the data collected from participants, we believe that by increasing the sample size and stimuli variation, we could hone in on what people are focusing on. Using eye-tracking has helped, and will help us, assess emotional engagement and provide objective data to complement subjective ratings.

6 CONCLUSION

This study investigated the capability of generative artificial intelligence (Gen AI) models to produce images that elicit specific emotional responses, such as calmness or unease. It assessed participants' reactions through eye-tracking metrics and self-reported emotional ratings. Our research indicates that Gen AI models can effectively generate images that provoke intentional emotional responses, as evidenced by consistent patterns in fixation durations, fixation counts, reaction times, and emotional ratings.

Cute images generated by the AI models resulted in longer reaction times, fixations across multiple AOIs, and high ratings on the cuteness scale. Based on the reaction time and fixations it could indicate that participants grew more relaxed. On the other hand, creepy images led to shorter reaction times, fixations on specific AOIs, and higher ratings on the scariness scale, indicating heightened emotional intensity.

However, Google and Meta's cute images introduced generation errors (Figure 2) that could have affected a participant'sresponses. Some participants rated these images as both cute and creepy, showing that the language models are not advanced, or trained, enough to complete the task as efficiently.

This study shows one use of how eye tracking technologies can capture emotional engagement and validate self-reports. Eye tracking provided insight into participants' mental strategies when viewing AI-generated images.

In conclusion, generative AI models have the capability to generate material to influence human emotion, but there is a way to go in training for AI to further this understanding. Eye tracking in this experiment showed the minute reactions in participants, which made this test unique and valuable. Future work should focus on improving Generative AI models to understand emotional impact more in-depth. Future work should also focus on analyzing eye tracking during moving stimuli and other forms.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation (grant IIS-1748380), and . . . Any opinions, findings, conclusions, or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsors.

REFERENCES

Cross, P. (2023). Artificial intelligence: A threat to middle-class, white-collar jobs? pages 10–11.

Varnum, M. E. W. (2023). Who's afraid of ai—and why? *Psychology Today*.

Yanardag, P., Obradovich, N., Cebrián, M., and Rahwan, I. (2021). Nightmare machine: A large-scale study to induce fear using artificial intelligence. In *International Conference on Innovative Computing and Cloud Computing*.