The Affective Nature of AI-Generated News Images: Impact on Visual Journalism

Sejin Paik[1], Sarah Bonna[1], Ekaterina Novozhilova[1], Ge Gao[1], Jongin Kim[1], Derry Wijaya[1,2], Margrit Betke[1]
[1]*Boston University,* [2]*Monash University Indonesia*
(sejin, sbonna, ekaterin, ggao02, jongin, wijaya, betke)@bu.edu

***Abstract*—This study explores the affective responses and newsworthiness perceptions of generative AI for visual journalism. While generative AI offers advantages for newsrooms in terms of producing unique images and cutting costs, the potential misuse of AI-generated news images is a cause for concern. For our study, we designed a 3-part news image codebook for affect-labeling news images based on journalism ethics and photography guidelines. We collected 200 news headlines and images retrieved from a variety of U.S. news sources on the topics of gun violence and climate change, generated corresponding news images from DALL-E 2 and asked study participants to annotate their emotional responses to the human-selected and AI-generated news images following the codebook. We also examined the impact of modality on emotions by measuring the effects of visual and textual modalities on emotional responses. The findings of this study provide insights into the quality and emotional impact of generative news images produced by humans and AI. Further, results of this work can be useful in developing technical guidelines as well as policy measures for the ethical use of generative AI systems in journalistic production. The codebook, images and annotations are made publicly available to facilitate future research in affective computing, specifically tailored to civic and public-interest journalism.**

***Index Terms*—Generative AI, Emotion analysis, Ethics of affective computing, Affective image generation, AI safety, Journalism ethics**

I. Introduction

In recent years, text-to-image generative AI systems based on multimodal deep learning models have shown advancements in their ability to produce affective images that are of comparable quality to human photographers and artists [1], [2], [3]. For example, Cosmopolitan was one of the first mainstream U.S. media publications to use an AI-generated cover art that depicted a completely novel and imaginative image of a female astronaut [4]. These generative AI systems are capable of producing images that are not only aesthetically compelling but also highly relevant to the specifications of human natural language inputs, making them an enticing tool in various content generation fields. However, the sociopolitical implications of these AI systems vary given the different goals and needs of the industry using the technology. While existing studies have examined the efficacy of generative AI systems and their affect-inducing capacities in the context of the arts and various creative fields [5], [6], [7], not many have investigated these systems' application within the public-interest, civic technology domains. Thus, this study focuses on a generative AI system, DALL-E 2 [8], in the context of visual journalism and its impact on human emotions.

With the emergence of computational journalism in the past decade [9], newsrooms have been using natural language processing (NLP) and computer vision-based tools to aid various news editorial work. Of late, synthetic media, i.e., artificially generated and/or manipulated media [10], are widely applied by news organizations. The Associated Press uses NLP to scan social media feeds for news gathering and deploys automatic generation of story summaries [11]. Similarly, the Los Angeles Times uses Quakebot, an algorithm that automates the reporting of latest earthquake news [12]. Most recently, a right-wing political news site made headlines due to their use of Midjourney, an AI-image generator, to produce news images that combined real-life stock photography with illustrations [13]. Generative AI tools open up opportunities and advantages for newsrooms as AI-images are completely unique and newsrooms won't need to compete with other organizations to select original images. Further, these tools are particularly useful for small and midsize newsrooms that are under budget restraints to hire human photojournalists or editors [14]. With the recent release of ChatGPT 4 [15], which allows multimodal input and output refinement, news images can be tailored to a particular story with even higher accuracy and detail.

Notwithstanding the outlined advantages, the potential misuse and dangers of AI-generated news images are high, such as the dissemination of fake news [16], the spread of mis/disinformation [17], and the perpetuation of harmful stereotypes [18]. The recent AI-generated image of Donald Trump being arrested before the event actually took place went viral and gathered more than 5 million views on Twitter [19]. In another instance, a set of photorealistic AI-generated images that depicted a fake earthquake that hit the Pacific Northwest in 2001 got on the "front page" of Reddit [20]. These highly-realistic photos coupled with descriptive captions resembling a news article format made it difficult for Reddit users to discern truth versus fiction. The spread of manipulated or distorted news such as these incidents are now easily attainable with generative AI programs like DALL-E 2, Midjourney, and Stable Diffusion, all of which are capable of producing high-quality images indistinguishable from reality. In journalism, this fundamentally undermines news professionals' integrity as images and graphics are seen as objective and context-adding tools for conveying a news story [21].

Given the foreseeable influx of harm that may appear from the application of generative AI in news creation, our study sets forth journalism-specific parameters that can be used to investigate the quality and impact of generative news images compared to images produced by humans. Methodologically, we created a codebook for affect-labeling of news images based on a set of journalism ethics and photography guidelines. A codebook is a common method used in qualitative data analysis as a way to systematically identify and classify patterns in various types of multimodal content [22]. This codebook can be used for any current or future generative news image datasets. Our 3-part news image codebook contains the following sections:

1) Emotional response to news images and headlines: Dominant emotional responses to news images consumed with and without a headline.
2) Photojournalism ethics: Emotional responses toward news images containing different levels of photojournalism characteristics such as context, newsworthiness and impact.
3) Image characteristics: Number of individuals and objects, depth of field, image quality and sophistication.

Subsequently, we created an affect-labeled news image dataset that includes 200 original news headlines and images retrieved from a variety of U.S. news sources on the topics of gun violence and climate change. Using the existing news headlines written by human journalists as the textual input, we generated corresponding news images from DALL-E 2. We then asked annotators to follow the codebook to annotate their emotional responses to the human new images as well as the DALL-E 2-generated images. Our results highlight the differences and similarities in emotional responses elicited by human news images and AI generated images. We also measured the level of journalistic values conveyed through the images such as context, newsworthiness, and impact. Lastly, we investigated the effect of modality on emotions and whether the consumption of certain modalities increase variance in emotional responses.

II. Background

A. Generative text-to-image AI

Until recently, AI-generated images have typically been the product of deep neural networks that are based on the architecture of General Adversarial Networks (GANs) [23]. GANs use two neural networks, i.e., generative and discriminative networks, to create new data. The generator (decoder) produces an image as output, while the discriminator (encoder) scores its realism, hence the ability to produce authentic-looking images. The original GAN model [23] has been extended over the past years, yielding powerful models with a broad range of abilities, e.g., adding on details to existing art works or generating faces of nonexisting people [10].

In 2021, OpenAI introduced diffusion models that outperform GANs [24] and later presented DALL-E 2 and CLIP models [25] that leverage language and vision inputs (using text and image encoders) to produce visuals (using an image decoder). An impressive part about these models is that they not only have an ability to manipulate and rearrange objects but also create realistic figures or objects that do not exist in real life [26]. Unlike GANs, generative text-to-image AI systems like DALL-E 2 and CLIP-guided image generation models like MidJourney are powered by a class of machine learning models known as transformer-based language models [27]. Transformer-based language models are typically used to generate a series of visual tokens, which are then transformed into an image using an image decoder network [25]. The image decoder network maps the visual tokens to a corresponding image by predicting the pixel values of the image. By training the model on a large corpus of text and image pairs, the model learns to generate images that are semantically consistent with the corresponding textual input.

B. Impact of generative AI images on human emotions

The ability of AI-generated images to elicit emotional responses in humans depends on several factors, such as the quality of the generated image, the context in which the image is being presented, as well as the individual differences that affect the emotional interpretation of the image. Studies have shown that deep learning models for image generation can closely model facial expressions and evoke emotional responses in humans that are similar to those elicited by real-world images [28]. An MIT-Media-Lab-led project called Deep Empathy created AI-generated images using neural style transfer to depict how the aftermath of a war might look in North American and European cities. The intended outcome of this project was to cause people to feel more connected and empathetic toward victims of disasters [29]. A different study developed a climate change visualizer that would produce AI-generated images of cities affected by environmental changes such as floods, storms, and wildfire [30].

While the main goal of these projects may have been to use emotions for a social cause, individuals' emotional responses to AI-generated images may not always be consistent or predictable across individuals or contexts, such as climate change, war, and terrorism. Existing works show that the elicitation of an emotion such as empathy or the interpretation of a smile can vary by individuals, given their cultural backgrounds and demographic characteristics [31]. Further, it is currently unclear how these human traits and contexts are taken into consideration as AI-generated image systems create emotion-inducing images. In particular, as emotion recognition is crucial in understanding the impact of news content [25], more research is needed in this area.

C. The role of emotions in visual journalism

While the goal of photojournalism is to convey objective truth and provide context to a news story [32], the consumption of images, as a product of photography, is inherently a subjective human experience. The best photojournalist work can stand alone to tell a news story without any additional information or textual component. News images, compared to news texts, can be powerful conduits for inducing emotions for individuals to feel more connected to the news [25]. In journalistic work, eliciting emotion is a tool used by photojournalists to ignite greater news engagement from news audiences by balancing objective truth and subjective perception [33], [34].

At the root of photojournalism is the idea that photojournalists' role is to depict events and situations as they happen in real time and document exactly what occurs in front of their eyes. In journalism ethics, it is noted that images should never be distorted, staged, or manipulated, and this is what defines the clear distinction between photojournalism and art-related photography [32]. In other words, evoking a particular emotional response towards a sensitive and/or polarizing political topic can be highly-damaging to the integrity and interpretation of the news story.

As such, photojournalism has faced great tensions over these foundational journalism principles of objectivity and documentation of reality, ever since the rise of digital photography software that has advanced the production of highly-stimulating and engaging visuals [33]. With generative AI, the news industry may face greater issues surrounding the production and dissemination of manipulated or false news images ripe with stimulating emotions. Previous research shows that mis/disinformation produced in multimodal form (e.g., a combination of text, image and graphics) tends to increase perceived news credibility and engagement intentions towards the misleading content [35]. Further, studies demonstrate that individuals who rely more on their emotions over logical reasoning are more likely to believe in fake news [36].

III. Method

A. Data Description and Collection

Two datasets of news article headlines were used in generating the data for DALL-E 2. One was taken from an existing gun violence (GV) news dataset that contained news headlines [37] and images [38]–[40]. The second dataset on climate change (CC) news, we newly collected for the purpose of this study. We randomly selected 100 headlines from the GV and CC datasets each and ensured that there were no duplicates. To generate news images, we passed news headlines as text prompts to DALL-E 2 via its API, where one image per headline was generated. DALL-E 2's content policy mentions that we should not create images of public figures [41]. In some instances, when a user passes a public figure's name into DALL-E 2, the model will generate an image that looks similar to the descriptions or the likes of an existing public figure. Hence, when selecting news headlines, we avoided those with the names of public figures as much as possible. If a headline violated DALL-E 2's other content policies, such as highly-violent and/or political content, another headline was randomly-selected and passed to the DALL-E 2 API. This cycle repeated until 100 images were obtained for each GV and CC news category (see Figure 1 for examples).

B. Codebook

A set of 10 questions regarding 1) the emotion impression of actual images from articles and DALL-E 2-generated (AI) images (with and without textual context), 2) photojournalism ethics and 3) image characteristics, were formulated into a codebook. The codebook can be downloaded from *http://www.cs.bu.edu/faculty/betke/aiem/codebook-ACII2023*.

News headline	Human - News Image	AI - News Image
Jacksonville mass killing once again proves the left's gun control 'solution' is a fleeting illusion		
Climate change protests snarl DC traffic as bizarre scenes unfold in capital		

Fig. 1: Example of news headlines, corresponding human-selected news image and AI-generated images

For the first codebook section on emotions, there were 12 emotions that annotators could choose from: *anger, disapproval, fear, sadness, confusion, curiosity, realization, surprise, relief, approval, admiration and excitement,* labeled from 1 to 12. We selected these emotions among a list of 28 emotion categories described by Alon and Ko [42] that we deemed most relevant to the context of news consumption, such as fear and anger [43]. Further, the set of 12 emotions consists of four emotions from each of the three sentiment categories *positive* (relief, approval, admiration and excitement), *negative* (anger, disapproval, fear, sadness), and *neutral* (confusion, curiosity, realization, surprise), selected to uphold consistency in statistical analysis. In this first coding section, annotators were first tasked with looking at just the human-selected and AI-generated images without looking at the headline and asked to provide their immediate emotional reaction. They would then repeat this same process, but in conjunction with reading the headline.

The second codebook section contains three questions about journalistic quality parameters such as context, informativeness and impact of a news image. The questions and the definition of each parameter were obtained from the photojournalism ethical principles outlined by The New York Times. The New York Times was selected because it "has long been one of the most prestigious and highest-profile newspapers in the world," according to InfluenceWatch [44]. In essence, context was defined as how specific or tailored each image was to the headline. Informativeness was defined as whether an image was descriptive enough for the coder to grasp the topic of the accompanying news story without reading the headline. Impact referred to whether an image when seen with the headline added more emotional weight to the news headline [32].

The topic of the last section on image characteristics was the sophistication of the image composition, i.e., how many objects or individuals were depicted and whether there was a clear focal point (defined as having focus on an object or individual with a blurred background) for both human-selected and AI-generated images. The codebook also included a sophistication comparison, i.e., which image among the human/AI pair felt more "sophisticated" or was of higher quality. Annotators were given the options of selecting human, AI, both, or neither.

A group of seven human annotators used the codebook to annotate 200 news items. The seven annotators were diverse in age, gender and ethnicity. For each news item, an annotator was given a data point that included the written news headline, a corresponding human-selected news image and the AI-generated image from the given headline. All news image data for annotators were labeled as human-selected and AI-generated. Prior to the actual annotation task, all seven annotators were briefed on the codebook and had two pre-testing rounds where annotators reviewed a small subset of the data and was given time to discuss their answer choices. After everyone was aligned on each question in the codebook, each annotator completed 29 data points, on average.

IV. Results

Descriptive statistical analysis was performed based on the annotations, and the following questions were focused on:

1) Emotion and sentiment analysis,
2) Impact of photojournalism characteristics on emotional responses,
3) Image characteristics.

A. Emotion and sentiment analysis

1) Multimodality effect on emotional response to news images: The most prevalent six emotions evoked by human-selected images (Fig. 2 top) before reading the headline were curiosity, fear, sadness, admiration, approval, and confusion, covering all three sentiment groups. After reading the headlines, annotators mostly reported positive emotions of approval and admiration, negative emotions of sadness, anger, and fear, but few neutral emotions. When seeing AI-generated images (Fig. 2 bottom) without the headlines, annotators' emotions were mostly confused or curious. After reading the headlines, participants reported a diverse range of emotions, such as approval, sadness, and anger. Nonetheless, 10% of the images were marked confusing versus 2.5% for human-selected images. Once the headlines were revealed, for both human-selected and AI-generated news images, the predominant emotions were approval and anger for the gun violence topic, and approval and sadness for the climate change topic.

2) Emotion and sentiment change: The emotion change annotators experienced between before and after reading the headline for human-selected and AI-generated images, respectively, are shown in Fig. 3. To quantify the change in emotion, we computed the absolute difference in emotion labels and tabulated them into four categories: 0: No change, 1–3: Slight change, 4–7: Moderate, and 8-11: Extreme. Based on this metric, slight or moderate shifts were observed for a majority

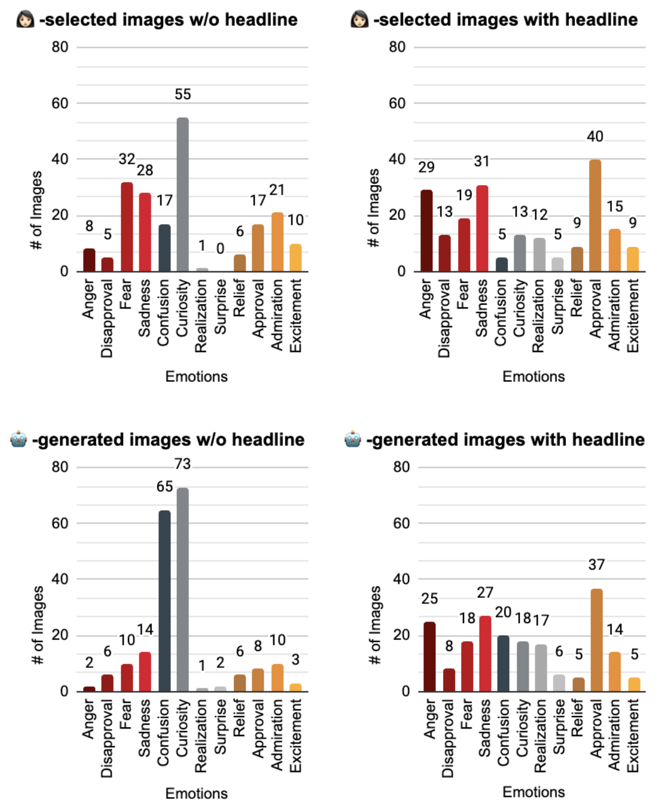

Fig. 2: Histograms of emotional responses to human-selected images before and after reading the corresponding headline (top) and AI-generated (bottom).

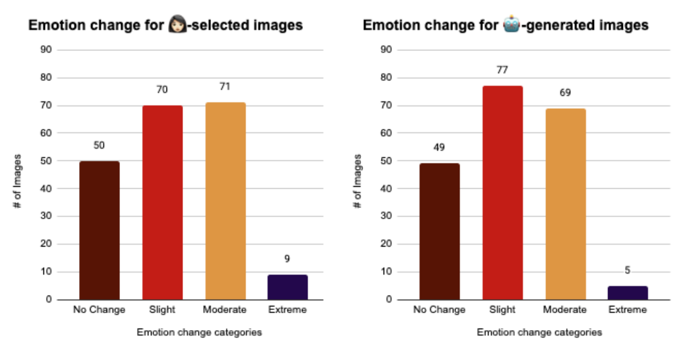

Fig. 3: Emotion change of human-selected and AI-generated images.

of the data. Significant differences in emotion change between the gun violence and climate change topics were not found.

We report the change of sentiment experienced by the annotator after the headline was revealed, as a change in the emotion groups *negative (2), neutral (1), and positive (0)*. To quantify the change in sentiment, we used the following metric: 0: No change, 1: Moderate, and 2: Extreme. For both gun violence and climate change topics, sentiment changes happened for about half the combined data (human-selected and AI-generated for each topic), i.e., 110/200=55% for gun violence combined data, and 99/200=49.5% for climate change combined data.

Emotion from 🤖 image-only	**Emotion from 🤖 image + headline**
Admiration	Fear
	 Rising emissions could drain foods like rice and wheat of their nutrients, causing a slow-moving global food crisis
Emotion from 🧑 image-only	**Emotion from 🧑 image + headline**
Admiration	Fear
	 Dems say GOP focus on mental health is redirection from gun control

TABLE I: After reading the headline, a participant could change their original emotion when only seeing the image. This applies to both AI-generated (top) and human-selected (bottom) images as well as news story topic, climate change (top) and gun violence (bottom). In this example, the human chose a different angle, producing two different negative emotions, while the AI connects the image with the headline better, yielding better consistency but not the same sentiment.

Looking at the overall data (climate change and gun violence together), there were 19 more human-selected images in the Extreme category than AI-generated ones. Taking all the sentiment change categories into consideration, for human-selected images, negative sentiment before and after reading the headline was the most common (49/200=24.5%), followed by positive before and after (31/200=15.5%) and neutral before to negative after (30/200=15%). For AI-generated images, neutral sentiment before and after reading the headline was the most common (53/200=26.5%), followed by switches from neutral before to negative or positive sentiments after (44/200=22% each). We show two examples of data with an emotion change in Table I.

3) Distribution of emotions of AI-generated images that were of high quality: There were 76 high quality AI-generated images, which were categorized as those AI-generated images that the annotators considered to be equally or more sophisticated than the human-selected ones, i.e. options “AI” or “Both” were selected by annotators (Fig. 4). Of the 76 images, 23 (30%) were solely AI-generated images (“AI” only selected). Figure 4 shows that among the AI-generated images that were of high quality, the top three emotions before reading the headline were curiosity (27/76=35.5%), confusion (12/76=15.8%) and sadness (9/76=11.8%).

Among 141 AI-generated images that evoked a neutral sentiment before reading the headlines, only 40 (40/141=28.4%) were considered to be of high quality. We further analyzed the quality differences of the top two neutral emotions: curiosity and confusion. First, out of the 73 AI-generated images that evoked curiosity before annotators read the headline, only 27 (27/73=37%) were considered to be high quality, while

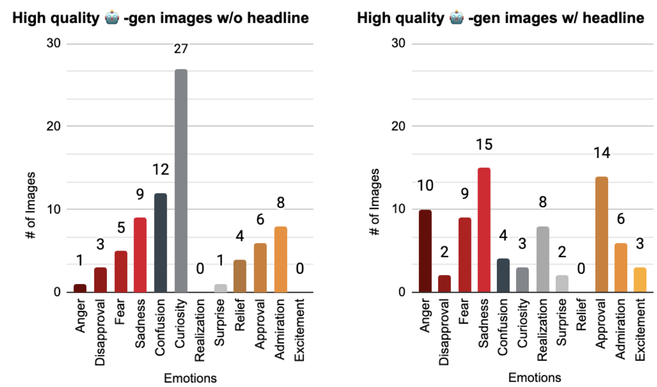

Fig. 4: Distribution of emotions of high quality AI-generated (right) images without and with the headline.

the remaining were considered low quality (46/73=63%). For confusion, out of the 65 AI-generated images before annotators read the headline, only 12 (12/65=18.5%) were considered to be high quality, while the remaining were considered low quality (53/65=81.5%).

Among the images that were considered high quality after reading the headlines, 36 images were considered negative (47.4%), 23 positive (30.3%), and 17 (22.4%) neutral. The top 3 emotions considered high quality after reading the headlines were sadness (15/76=19.7%), approval (14/76=18.4%) and anger (10/76=13.2%). For the neutral emotions, there were 18 AI-generated images that evoked curiosity, with only 3 (3/18=16.7%) of the images considered as high quality, while the remaining as low quality. For confusion, there were 20 AI-generated images with only 4 (4/20=20%) of the images considered as high quality.

B. Impact of photojournalism characteristics on emotional responses

Photojournalism comprises of the context, informativeness, and impact of the image. The annotators’ provided their opinions on these three photojournalism characteristics towards both the human-selected and AI-generated images. We found that annotators considered both human-selected and AI-generated images to be well-tailored to the headline (i.e. having a lot of context), with a large number of the images in the “Very much” or “Much” categories (human 117/200, AI 82/200) with AI-generated images lagging significantly by 17.5% points. Human selection was more successful than AI generation in ensuring that the images were informative (human 91/200, AI 46/200), with AI-generated images lagging by 22.5% points. Images were considered to have impact (human 109/200, AI 74/200), with AI-generated images lagging by 17.5% points. Among the three desirable properties of a news image, providing context, being informative, and having impact, our data reveals that being informative was the most difficult to accomplish, particularly for the AI. Among the images deemed to be highly informative, the prevalent emotion evoked was curiosity, an unexpected combination.

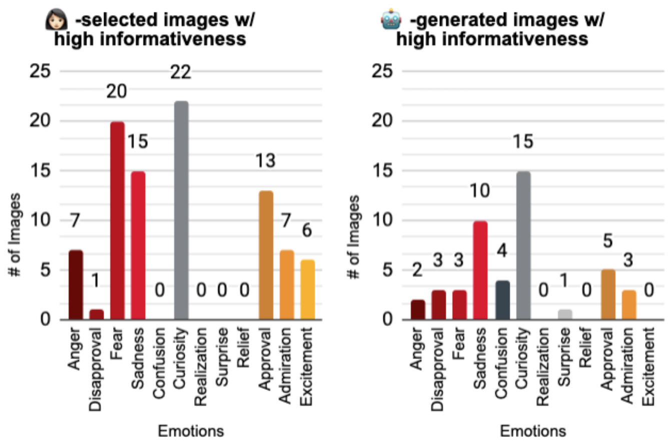

Fig. 5: Distribution of emotions of human-selected (left) and AI-generated (right) images where the level of informativeness was high.

1) Distribution of emotions for human-selected and AI-generated images where the clarity of context was high: Images with high clarity of context are defined as images where the context is in the "very much" or "much" categories. Context considers how tailored the image is to the headline, so the emotion evoked by the image after annotators read the headline is considered. There were 117 human-selected (58.5%) and 82 AI-generated (41%) images with high clarity of context.

The five most prevalent emotions evoked by human-selected images with high clarity of context were approval (18.8%), sadness (15.4%), anger (13.7%), fear (12%) and admiration (8.5%). For AI-generated images with high clarity of context, the five most prevalent emotions were approval (19.5%), sadness (15.9%), fear (12.2%), followed by a tie between admiration (11%) and anger (11%).

Only a small number of images in the neutral sentiment group were considered to provide high context clarity (human 19/117, AI 17/82). Almost half of the human-selected images with high context clarity were negative (53/117=45%). There were fewer human-selected, high context images with positive sentiment (45/117=38%). The AI-generated images annotated as high context clarity also evoked more negative (35/82=42%) than positive sentiments (30/82=37%).

2) Distribution of emotions for human-selected and AI-generated images with a high level of informativeness: Images with a high level of informativeness are defined as being in the "Definitely" or "Probably" informative categories. Informativeness considers whether the image alone is sufficiently descriptive for a person to grasp that it is the lead to a news story. The emotion evoked by the image before annotators read the headline was considered. There were 91 human-selected (45.5%) and 46 AI-generated (23%) images with high levels of informativeness. The four most prevalent emotions evoked by human-selected images with high level of informativeness were curiosity, fear, sadness, and approval (Fig. 5), with negative emotions prevailing (43/91=47%). For AI-generated images with high level of informativeness, the four most prevalent emotions were curiosity, sadness, approval and confusion.

3) Distribution of emotions for human-selected and AI-generated images where the level of impact was considered to be high: Images with a high level of impact are defined as images deemed in the "very much" or "much" impact categories. Impact considers whether the image added more emotional weight to the headline, so the emotion evoked by the image after annotators read the headline is considered. There were 109 human-selected (54.5%) and 74 AI-generated (37%) images with high levels of impact. The four most prevalent emotions evoked by human-selected images with high impact were approval (18.3%), sadness (17.4%), anger (15.6%), and fear (12%), with negative emotions prevailing (55/109=50.4%). For AI-generated images with high impact, the four most prevalent emotions were anger (20.3%), sadness (17.6%), approval (16.2%) and fear (13.5%), with negative emotions prevailing (40/74=54%).

C. *Image characteristics*

1) Depth of Field: Most images have a clear focal point and background, i.e., large depth of field (Human 85/200 and AI 83/200). This is followed by images with a blurry background, i.e., small depth of field (human 57/200, AI 48/200).

2) Number of objects/individuals: Most images focus on a single object or person (human 70/200, AI 88/200), followed by the images with 2–4 objects/individuals (human 55/200, AI 58/200), 5–10 objects/individuals (human 39/200, AI 30/200), and larger groups (human 36/200, AI 24/200). For a given headline, we compared the human choice of showing a certain number of objects or individuals with the choice of the AI. Interestingly, the majority of the image pairs have the same (80/200=40%) or very similar (74/200=37%) number of objects/individuals.

3) Human versus AI image news quality: In general, the human-selected images were of higher quality than AI-generated ones. Out of 200 data points, 110 human-selected images were deemed to be of higher quality, compared to 23 AI-generated ones. There were 53 cases where both human-selected and AI-generated images were deemed to be of high quality, and 14 cases where neither of them were thought to be of high quality.

V. Limitations of AI model and experiment

Our results show that DALLE-2 was able to capture some context extracted from the news headline, but still lacks in its technical capacity to provide journalistic values of informativeness and impact. This can be seen from the results described in sections IV B.2 and B.3.

However, it is possible that there were potential confounding and biased responses elicited from the design of our annotation procedure such as providing labels on news images created by the AI and selected by humans. In future works that build on our codebook, a blind test with randomization of question-answering order could provide deeper insights into the quality

Fig. 6: Depiction of objects and figures. Human (left) and AI (right) provided images. 1) & 2) News images of 3D-printed guns are common. AI instead shows the 3D printing process (row 1) or other context (row 2, headline "Chicago, suburban libraries brace for the question: Can I print out a 3D gun?" 3) AI generates a fake, yet stereotypical and real-looking person when the headline contains phrasing such "tech mogul" and/or "CEO." These phrases elicited a male figure, with a hand gesture similar to Steve Jobs' gestures, and the camera angle tilted upward.

perception and affective responses towards AI-generated news images.

While we acknowledge that at the time of data collection, there were more sophisticated visual generative AI models (e.g. Midjourney and Stable diffusion), we used DALLE-2 as a case study to examine the average capacity of the image generative AI models that publicly exist today. The goal of our study was not to examine how DALLE-2 could generate "better" emotion-driven news images. Rather, we wanted to assess what types of news images the AI model could generate without any detailed prompting and assistance around emotion cues. Our prompts simply asked DALLE-2 to generate an appropriate news image for the textual news headline provided. Future work can explore a larger dataset that includes news images generated by multiple image generative tools and explores emotion prompting.

VI. Discussion

There is an increasing opportunity for AI-generated image systems to become highly-resourceful tools for assisting and even automating news production in the journalism industry. However, this technology also poses challenges for news professionals as they strive to uphold established journalistic principles of transparency, objectivity, and efforts to minimize harm. The first issue with AI-generated news images stems from one of the main photojournalism ethical principles which states that no real-life images should be distorted, manipulated, stereotyped, or staged [32]. As such, the use of generative AI to produce hyper-realistic images fundamentally goes against the ethics of photojournalism.

Results from our emotion-labeled dataset of human and AI news images show that the distribution of emotions for human-selected and AI-generated images marked as having high

Fig. 7: AI-generated images that caused confusion (left two) or curiosity (right two) when seen without the headline. The two confusion-causing images lack clarity and quality (the AI model seems to try to depict an insect and politician, respectively) while the two curiosity-causing images portray contexts that are visually compelling (drinking birds, white male in uniform during interview).

clarity of context are similar, suggesting that those human-selected and AI-generated images that are well-tailored to the headlines evoke similar emotions. This foreshadows that generative AI systems have the potential to produce images at the quality of a human photojournalist and would be able to trigger similar emotional responses to those that were taken and selected by human photojournalists. This can cause serious harm to the integrity of journalism when a news image is generated for a completely fake event or story line, as the news audience would not be able to discern the truth. When the headline contained general descriptions or words that indicated a certain type of person (e.g. pertaining to one's health: "a mentally distressed person" or a person's job title or ranking), we found instances of stereotypical caricatures based on existing societal gendered or racial norms (Fig. 6.3).

A second issue with AI-generated images is that they could elicit a more diverse and/or intense range of emotional responses from audiences, making it difficult for human journalists to maintain control over their intended narrative. Our results show that the distribution of emotions for AI-generated images with a high level of informativeness was more diverse than human-selected ones. The AI-generated images elicited 9 emotions (human-selected only 8), and there was a more balanced distribution of AI-generated images across the 9 emotions than for the human-selected images. Most human-selected images with a high level of informativeness evoked 38% of fear and sadness, both stemming from the negative sentiment category. In contrast, most AI-generated ones evoked curiosity, sadness and approval which spanned positive, neutral, and negative sentiments.

When further comparing the emotional response differences evoked by human and AI news images, we found that 69% of AI-generated images before the annotators read the headline noted emotions of curiosity and confusion (Fig. 7). This shows that AI-generated images do not seem to be able to capture the main idea of the headline clearly, leading to such ambiguous emotions when looking at the images alone. On the other hand, the spread of emotions annotated for AI news

Fig. 8: Human (left) and AI (right) provided images. 1) AI generated more graphical content compared to human images. 2) Protest images provided by Human and AI are similar (here, the US Capitol appears in the background of the AI-generated image, providing context that the photograph does not). 3) AI is accurate in generating images of landscapes and disasters, here a methane leak is reported.

images after reading the headline was a lot more diverse. This shows that the emotions evoked after reading the headlines are heavily influenced by the headlines themselves, speaking to the interaction between textual and visual modalities.

When comparing the emotional impact of news images in different news topics, we found that the AI system was better at generating both higher news quality and human-like images in the context of climate change compared to gun violence news. For example, the climate change headlines we used for prompting the AI images often contained depictions of wildlife, landscape, and natural disasters (Fig. 8). Climate change news also contained a lot of protest-related headlines, and in these instances, we found that AI-generated images tended to provide comparable or more context to the news story than did the human-selected images.

Furthermore, as gun violence news headlines contained high levels of politician names or triggering phrasings (e.g., *murdered, killed,* etc.), the AI-generated news images were more varied in their depictions of the headline compared to the human-selected news images. In cases where headlines contained words related to guns as an object, the AI system avoided generating a clear depiction of the weaponry, causing confusion, for example, when participants saw an image of a 3D printing process (Fig. 6-1, AI). In addition, regardless of the news topic, the AI model generated more image visuals that resembled graphic figures, posters and design (Fig. 8 top right). Future research with a larger dataset could better understand what textual elements trigger a graphical visual over a realistic image.

In conclusion, our findings highlight the need for caution and transparency when using AI-generated news images in journalism, as they have the potential to influence audience perceptions and reactions in ways that may not align with journalistic principles of upholding integrity and truth towards the people and events in a given story. This could have significant implications for the perceived credibility of news and individuals' willingness to engage in civic activities. For practical implications, understanding the differences between emotional responses to human-selected and AI-generated news images can inform the development and use of AI systems in journalism. This study is intended for anyone in the journalism field as well as those developing and governing generative AI technologies. For example, if AI-generated news images continue to elicit significantly diverse and inconsistent emotional responses compared to human-selected ones, it may indicate a need for more transparency and oversight in the use of AI in news reporting. Using journalism as a case for technology used for civic engagement and public affairs, our work aims to highlight the need to move beyond highly generalized ethical frameworks for affective AI systems and to move towards these systems becoming more aligned to industry- and domain-specific values.

VII. Ethical Impact Statement

This study involved both the creation of an emotion analysis codebook and the work of human annotators to annotate their emotional responses to human-selected versus AI-generated news images from existing news headlines. We provided clear instructions on how annotators could answer the questions and asked for their consent to participate in the annotation task. IRB review resulted in exemption due to minimal risk to the annotators. The study certainly raises issues related to potential negative societal impact, particularly in the use of AI-generated images in news reporting. Such generative images may have the potential to reinforce biases and stereotype which can cause serious harm to individuals. The publicly-accessible generative tools can perpetuate mis- and dis-information and contribute to the spread of fake news. To mitigate this risk, academic scholars and practitioners need to carefully consider the implications when it comes to the application of AI-generated images for civic-interest news and to ensure that they are representative of a diverse range of perspectives and voices. Finally, the study raises issues related to its generalizability. The study's findings as well as its limitations, such as the news dataset size, limited exploration of topics, and inherent biases from the human annotators themselves, may not be generalizable to other cultural backgrounds, countries and contexts.

VIII. Acknowledgements

This work is supported in part by the U.S. NSF grant 1838193 and DARPA HR001118S0044 (the LwLL program). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes. The views and conclusions contained in this publication are those of the authors and should not be interpreted as representing official policies or endorsements of NSF, DARPA, or the U.S. Government.

References

[1] Y. Wang, S. Shen, and B. Y. Lim, "RePrompt: automatic prompt editing to refine AI-generative art towards precise expressions," *arXiv:2302.09466*, 2023.

[2] W. Chen, H. Hu, C. Saharia, and W. W. Cohen, "Re-imagen: retrieval-augmented text-to-image generator," *arXiv:2209.14491*, 2022.

[3] H.-K. Ko, G. Park, H. Jeon, J. Jo, J. Kim, and J. Seo, "Large-scale text-to-image generation models for visual artists' creative works," in *Proceedings of the 28th Int. Conf. on Intelligent User Interfaces*, 2023, pp. 919–933.

[4] "Dall-E 2 made its first magazine cover," 06 2022. [Online]. Available: https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/

[5] T. Galanos, A. Liapis, and G. N. Yannakakis, "AffectGAN: affect-based generative art driven by semantics," in *9th Int. Conf. Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2021.

[6] S. Salevati and S. DiPaola, "A creative artificial intelligence system to investigate user experience, affect, emotion and creativity," *Electronic Visualisation and the Arts (EVA 2015)*, pp. 140–147, 2015.

[7] D. Jha, H. Chang, and M. Elhoseiny, "Wölfflin's affective generative analysis for visual art." in *ICCC*, 2021, pp. 429–433.

[8] "Dall·E 2." [Online]. Available: https://openai.com/product/dall-e-2

[9] C. Anderson, "Towards a sociology of computational and algorithmic journalism," *New Media & Society:15(7)*, pp. 1005—1021, 2012.

[10] L. Whittaker, T. C. Kietzmann, J. Kietzmann, and A. Dabirian, ""All around me are synthetic faces": the mad world of AI-generated media," *IT Professional:22(5)*, pp. 90–99, 2020.

[11] "Artificial Intelligence: AP." [Online]. Available: https://www.ap.org/discover/artificial-intelligence

[12] "What is the quakebot and how does it work?" 05 2019. [Online]. Available: https://www.latimes.com/la-me-quakebot-faq-20190517-story.html

[13] J. Vincent, "This US political site has started using AI to illustrate its articles," 09 2022. [Online]. Available: https://www.theverge.com/2022/9/20/23363073/ai-art-generator-illustrator-news-site-the-bulwark

[14] "What AI image generators can do for newsrooms," 10 2022. [Online]. Available: https://thefix.media/2022/10/20/what-ai-image-generators-can-do-for-newsrooms

[15] OpenAI, "GPT-4 technical report," 2023.

[16] "ChatGPT is making up fake guardian articles. here's how we're responding — Chris Moran," 04 2023. [Online]. Available: https://www.theguardian.com/commentisfree/2023/apr/06/ai-chatgpt-guardian-technology-risks-fake-article

[17] T. Thomson, D. Angus, P. Dootson, E. Hurcombe, and A. Smith, "Visual mis/disinformation in journalism and public communications: Current verification practices, challenges, and future opportunities," *Journalism Practice:16(5)*, pp. 938–962, 2022.

[18] F. Bianchi, P. Kalluri, E. Durmus, F. Ladhak, M. Cheng, D. Nozza, T. Hashimoto, D. Jurafsky, J. Zou, and A. Caliskan, "Easily accessible text-to-image generation amplifies demographic stereotypes at large scale," *arXiv:2211.03759*, 2022.

[19] C. Warzel, "People aren't falling for AI Trump photos (yet)," 03 2023. [Online]. Available: https://www.theatlantic.com/technology/archive/2023/03/ai-generated-fake-trump-indictment-images/673513/

[20] M. Novak, "AI creates photo evidence of 2001 earthquake that never happened," 03 2023. [Online]. Available: https://www.forbes.com/sites/mattnovak/2023/03/27/ai-creates-photo-evidence-of-2001-earthquake-that-never-happened/?sh=590686f53985

[21] M. Carlson, "News algorithms, photojournalism and the assumption of mechanical objectivity in journalism," *Digital Journalism*, vol. 7, no. 8, pp. 1117–1133, 2019.

[22] K. M. MacQueen, E. McLellan, K. Kay, and B. Milstein, "Codebook development for team-based qualitative analysis," *Cam Journal:10(2)*, pp. 31–36, 1998.

[23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 1–9.

[24] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Advances in Neural Information Processing Systems:34*, pp. 8780–8794, 2021.

[25] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," *arXiv:2204.06125*, 2022.

[26] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. B. D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, F.-F. Li *et al.*, "On the opportunities and risks of foundation models," *arXiv:2108.07258*, pp. 1–214, 2021.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv:1706.03762*, 2017.

[28] K. Weitz, T. Hassan, U. Schmid, and J.-U. Garbas, "Deep-learned faces of pain and emotions: elucidating the differences of facial expressions with the help of explainable AI methods," *tm - Technisches Messen:86*, 06 2019.

[29] "Deep empathy by MIT Media Lab," 2017. [Online]. Available: https://deepempathy.mit.edu/

[30] A. Luccioni, V. Schmidt, V. Vardanyan, and Y. Bengio, "Using artificial intelligence to visualize the impacts of climate change," *IEEE Computer Graphics and Applications:41(1)*, pp. 8–14, 2021.

[31] Jenka, "AI and the American smile," 03 2023. [Online]. Available: https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf

[32] "What is photojournalism and why is it important?" [Online]. Available: https://nytlicensing.com/latest/marketing/what-is-photojournalism/

[33] C. Beckett and M. Deuze, "On the role of emotion in the future of journalism," *Social Media + Society:2*, 09 2016.

[34] "Objective truth vs. subjective perception." [Online]. Available: https://art21.org/read/objective-truth-vs-subjective-perception/

[35] J. Lee and S. Y. Shin, "Something that they never said: multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news," *Media Psychology:25(4)*, pp. 531–546, 2022.

[36] C. Martel, G. Pennycook, and D. G. Rand, "Reliance on emotion promotes belief in fake news," *Cognitive research: principles and implications:5*, pp. 1–20, 2020.

[37] S. Liu, L. Guo, K. Mays, M. Betke, and D. T. Wijaya, "Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence," in *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*, 2019, pp. 504–514.

[38] C. Reardon, S. Paik, G. Gao, M. Parekh, Y. Zhao, L. Guo, M. Betke, and D. T. Wijaya, "BU-NEmo: an affective dataset of gun violence news," in *13th Language Resources and Evaluation Conf.*, 2022, pp. 2507–2516.

[39] G. Gao, S. Paik, C. Reardon, Y. Zhao, L. Guo, P. Ishwar, M. Betke, and D. T. Wijaya, "Prediction of people's emotional response towards multi-modal news," in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, 2022, pp. 364–374.

[40] I. C. Tourni, L. Guo, T. H. Daryanto, F. Zhafransyah, E. E. Halim, M. Jalal, B. Chen, S.-C. Lai, H. Hu, M. Betke, P. Ishwar, and D. Wijaya, "Detecting frames in news headlines and lead images in U.S. gun violence coverage," in *Conference on Empirical Methods in Natural Language Processing*, Nov. 2021, pp. 4037–4050.

[41] "Content policy." [Online]. Available: https://labs.openai.com/policies/content-policy

[42] D. Alon and J. Ko, "Goemotions: A dataset for fine-grained emotion classification," 10 2021. [Online]. Available: https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html

[43] F. Aslam, T. M. Awan, J. H. Syed, A. Kashif, and M. Parveen, "Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak," *Humanities and Social Sciences Communications:7(1)*, 2020.

[44] "Influencewatch," 07 2023. [Online]. Available: https://www.influencewatch.org/for-profit/new-york-times/