

Human Recognition of LLM-Guided, Emotion-Driven Volume Renderings

Maximilian Jellen*

TU Wien, Austria

Renata G. Raidou†

TU Wien, Austria

ABSTRACT

We investigate whether large language models (LLMs) can generate volume renderings whose intended emotions are reliably recognized by humans. Using GPT-4o, we assign visual parameters (object color, background, and lighting) to neutral volumetric data based on literature-derived emotion and sentiment mappings. We render 28 LLM-guided, emotion-driven scenes in Unity and evaluate in a user study with 44 participants whether the emotions and sentiments for each image are recognized correctly. Preliminary results indicate that overall sentiment (positive vs. negative) is reliably conveyed, while precise emotion recognition (e.g., sadness, fear, anger, etc.) remains more challenging. Negative emotions such as anger were consistently identified, whereas surprise, fear, and some depictions of sadness were often misinterpreted. Entropy analysis revealed the most ambiguous images and highlighted cases that lead to inconsistent recognitions. Our study showcases the potential of LLMs for scalable, emotion-aware 3D scene generation and provides insights into human perception of LLM-guided visual emotional cues.

Index Terms: Large Language Models (LLMs), Emotion Recognition, Human Perception, Affective Computing, Rendering.

1 INTRODUCTION

Conveying emotions through visual media is a central challenge in psychology, computer graphics, and affective computing, with applications in virtual reality, human-computer interaction, and adaptive storytelling. Traditional methods rely on fixed mappings between low-level visual features, e.g., color, brightness, and contrast, and specific emotions or sentiment categories [1, 5, 8]. While effective, these approaches are labor-intensive and difficult to scale across diverse objects and scenes. Recent advances in large language models (LLMs) offer a promising alternative, but their potential for emotion-driven 3D scene generation remains unexplored. Open questions include whether LLMs can generate parameters that *accurately reflect the intended emotions* (i.e., in agreement with literature) and, most importantly, whether humans can *reliably recognize* these generated cues (i.e., whether generated cues align with human expectations).

In this work, we explore the capabilities of GPT-4o to assign object colors, background hues, and lighting intensities that reflect target emotions (e.g., happiness, sadness, etc.) and sentiments (positive vs. negative) in volume renderings. Using literature-based mappings for the emotions and the sentiments, we develop and validate an LLM-guided approach to generate visual parameters, which we use to render 28 scenes with neutral objects. Finally, in a user study with 44 participants we assess human recognition of the generated emotions and sentiments.

Our contributions are twofold: (*i*) we introduce a framework based on LLMs for emotion-driven 3D scene generation, validated against established literature-based visual emotion mappings, and (*ii*) we provide empirical evidence on human perception of these LLM-generated emotional cues. Our study showcases initial insights

into both the potential and limitations of the proposed LLM-guided visual adjustments in volume renderings.

2 RELATED WORK

The study of emotion classification has deep roots in psychology and behavioral sciences, with applications extending into computer science, particularly in affective computing, computer vision, and user experience design. Accurate emotion categorization is essential for systems aiming to understand or predict human emotional responses. Various models serve as the foundation for computational approaches, though there is no universally accepted model [7, 13].

Emotion models can be broadly divided into dimensional and categorical frameworks. Dimensional models, such as the *Pleasure-Arousal-Dominance (PAD)* model, represent emotions along continuous axes using three dimensions: pleasure (affective valence), arousal (intensity of activation), and dominance (perceived control) [11, 12]. While these models allow quantification of the emotional states, they are known to abstract away from intuitively recognized emotional categories. Categorical models define discrete emotion labels, often grounded in observable behavior. Ekman's theory of *six basic emotions* (anger, disgust, fear, happiness, sadness, and surprise) is widely cited and forms a basis for many studies [2, 5, 9]. The *wheel of emotions* expands upon this framework, arranging eight primary emotions in a visual model that illustrates intensity and combinatory relationships [10].

Assigning emotions to visual properties has been extensively studied. Sutton et al. [8] established color-emotion associations via surveys. Analyses of social media imagery reveal correlations between low-level features (brightness, saturation, hue) and perceived sentiment [3, 5]. For example, cooler colors in Instagram images were linked to positive sentiment due to outdoor sky backgrounds, highlighting the complexity of generalizing emotion-color mappings [1]. *Applying emotions* to images involves modifying visual parameters according to defined rules. Wang et al. [14] applied color themes based on single-word inputs, while Peng et al. [9] adapted one image's emotion histogram to another using Ekman's emotions and valence-arousal values. Although no prior work uses LLMs explicitly, they have been recently shown to assist visual tasks, such as autonomous visualization and iterative feedback for rendering [4]. To the best of our knowledge, applying emotions in 3D scenes and controlling their visual parameters through LLMs to affect the perceived emotions remains unexplored.

3 METHODOLOGY

The aim of our work was to evaluate whether LLMs can effectively apply emotions to a 3D scene by adjusting visual parameters in a way that humans can accurately perceive said emotions. Specifically, we ask *whether LLM-generated visual adjustments can convey intended emotions in volume renderings and how accurately human observers can recognize them*. In this manuscript, we use the following two concepts: *emotion*, referring to specific natural-language descriptors (e.g., sadness or happiness), and *sentiment*, representing broader categories of either positive or negative emotional valence.

Our workflow is depicted schematically in Figure 1. The workflow begins with selecting neutral 3D objects, and defining a set of “ground truth” target emotions and sentiment categories based on literature (Figure 1, cyan). The emotions and sentiment categories are

*e-mail: e12024736@student.tuwien.ac.at

†e-mail: mail@renataraidou.com

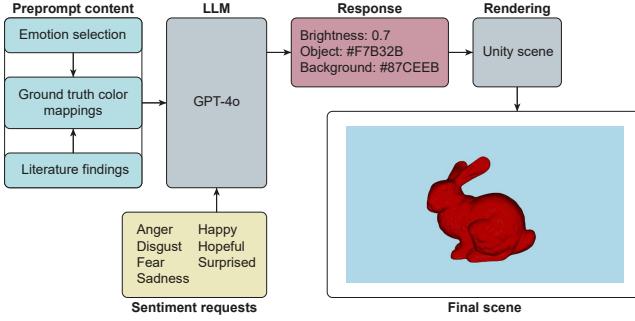


Figure 1: LLM-guided, emotion-driven rendering workflow: Emotions and sentiments are mapped to visual parameters using prior literature-based rules. This is used as a prompt for GPT-4o to generate a set of visual parameters for volumetric 3D rendering in Unity.

employed as the preprompting content for LLM-generated parameters for object color, background color, and light intensity (Figure 1, magenta) based on a sentiment request (Figure 1, yellow). The LLM response entails a set of visual parameters applied to volumetric 3D scenes through Unity. The resulting renderings are then presented in a user study, where participants select the perceived emotion and sentiment for each image. Finally, the collected data allow us to assess the effectiveness of the LLM in applying emotions to the renderings and to evaluate the accuracy of emotion and sentiment recognition by humans.

Preprompt Content Existing research informed the selection of a set of emotions, visual parameters, and ground-truth rules connecting image properties to emotions and sentiment. The selected emotions included Ekman’s six basic emotions: anger, disgust, fear, sadness, happiness, and surprise, along with one additional emotion: *hopeful*. The inclusion of *hopeful* targeted a more balanced distribution between positive and negative sentiments and provided an additional unique set of associated colors, identified by Sutton and Altariba [8]. Similar selection methods have been used in previous studies linking emotions to image features, such as by Chen et al. [9] and Machajdik and Hanbury [5]. For sentiment analysis, the emotions were grouped into a *positive* (happy, hopeful, surprised) vs. a *negative* (anger, disgust, fear, sadness) sentiment category.

Ground-truth mappings between emotions/sentiments and visual parameters have been previously derived by Sutton and Altariba [8] and Amencherla and Varshney [11]. The former work provides color frequencies based on survey data, with the top 95% of choices per emotion being included in the ground-truth rules. The latter work provides insights into the interaction of brightness and background color with sentiment, supporting a systematic mapping of visual parameters to emotions. Due to their explicit and grounded connections between low-level image features and emotions or sentiment, we also make use of these two types of mappings in our approach.

LLM Prompting and Response We chose GPT-4o to provide values for object color, background color, and light intensity parameters to be used while rendering our desired objects. Colors have been requested in hexadecimal format and lighting as a floating-point value between 0 and 1. Initial trials used the web interface, but later this approach was replaced by a Python script interacting directly with the API, enabling automation, better control over model versions, and reproducibility. Each request included a prompt containing detailed information about variable behavior and the aforementioned ground truth color mappings, following a heuristic-centric approach similar to Liu et al. [4]. Practically, the prompt instructed brightness levels, background, and object color

selection, and provided concrete color-emotion mappings based on prior surveys [8, 1]. The full prompt is included in the **Appendix**.

Volume Rendering with LLM-based Scene Parameters We employed four neutral objects from the [Open Scientific Visualization Datasets](#) by Klacansky: the Stanford bunny, the Boston teapot, the Christmas tree, and the Christmas present. These objects were selected for their presumed emotional neutrality and familiarity. For instance, we excluded human anatomical datasets to avoid unintended emotional bias and to ensure that participants were familiar with the depicted objects. 3D scenes were constructed in Unity, using the [UnityVolumeRendering](#) library to render volumetric data. Additional scripts ensured consistent camera positioning and automated screenshot capture. The previously LLM-generated response containing the visual parameters was manually applied, adjusting object colors via transfer functions, background color through camera settings, and light intensity within shader code.

4 RESULTS

4.1 LLM Effectiveness in Emotion-Driven Rendering

To ensure reliability, all 28 rendered scenes (4 objects \times 7 emotions) were systematically reviewed against the literature-based mappings provided in the prompt. The authors manually verified whether each image adhered to the expected associations between color, brightness, and emotion. This step served as a quality control, identifying cases where the model’s output diverged from the intended design. Out of the 28 images, only two were identified as inconsistent with the mappings (Figure 2): one prompted as anger was fitting better to surprise (a), and the other was fitting better to hopeful despite being prompted as sadness (b). The “better deemed fits” were based on background and object color cues. However, these two images were not excluded from the user study, as we were interested in examining how participants would respond to such cases.

4.2 Human Recognition of LLM-Generated Emotions

User Study Design and Execution To preliminarily assess how humans recognize LLM-generated emotions in the renderings, we conducted an online user study with 44 participants. The participants, recruited through personal networks, provided demographic information including age, gender, and field of study. The participants were aged 18–55 (mean age 23.8, median 23), with balanced gender representation (50% male, 45.5% female, 4.5% undisclosed) and diverse educational backgrounds. A screening question filtered out color vision deficiency respondents. Instructions emphasized quick, first-impression emotion selection for 28 images presented in a randomized order, disregarding the object’s inherent content.

Data Analysis After user study completion, the data were analyzed using RStudio. Data analysis focused on the correctness of



Figure 2: The two images found to be non-compliant with the ground truth rules for the requested emotion: (a) *Anger* was initially requested, but deemed to fit *surprise* better. In the user study, this was the most ambiguous image for emotion recognition. (b) *Sadness* was initially requested, but deemed to fit *hopeful* better. In the user study, this image was most frequently assigned to an emotion of a different sentiment category than intended.

sentiment and emotion assignments relative to the literature-based ground truth color mappings discussed above.

Across all trials, participants correctly identified the intended *emotion* in average in 37.7% of cases. *Anger* was the most accurately recognized emotion (51.7%), whereas *surprise* was the least (19.9%) (see Figure 3). However, when looking at the results per image, in 21 out of 28 images, the most frequently chosen emotion matched the one originally intended. In other words, even if individuals often disagreed, the group as a whole tended to lean toward the correct emotion for most images. For the *sentiments*, participants matched the intended positive or negative valence in average in 79.3% of cases, indicating that sentiment was generally clear. Correct assignments were higher for the negative (82.7%) than for positive ones (75.9%) (Figure 4; Chi-squared test for equality of proportions with $p < 0.001$), supporting prior literature that argues that negative emotions are more reliably conveyed [6]. Exact binomial tests confirmed that both sentiment and precise emotion assignments were significantly above chance ($p < 0.001$). There was no significant evidence for gender correlating with accuracy, as a Kruskal-Wallis test for differences in accuracy by gender produced $p = 0.38$ for emotion and $p = 0.24$ for sentiment. All assignment results are summarized in the confusion matrices of emotions (Figure 3) and sentiments (Figure 4) below.

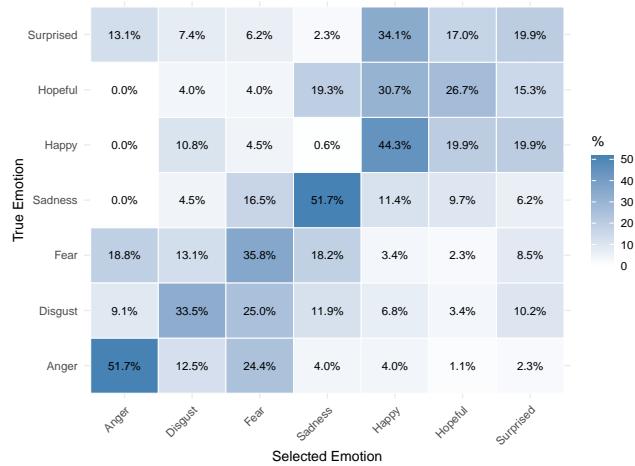


Figure 3: Confusion matrix of *emotion* recognitions per category: clusters in the 4x4 bottom left and 3x3 top right sections indicate that emotions are often confused with others of the same sentiment.

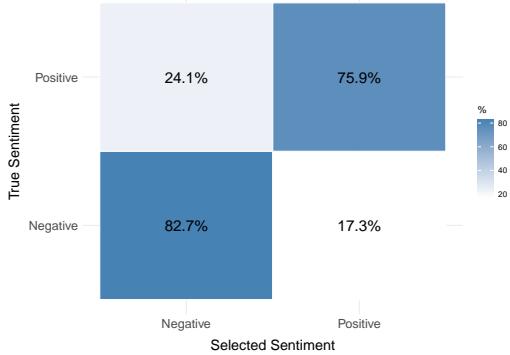


Figure 4: Confusion matrix of negative vs. positive *sentiment* recognition. Negative sentiments are more recognizable ($p < 0.001$).

Entropy Analysis Entropy analysis was conducted to assess ambiguity in the LLM-generated images. High entropy results indicate images with evenly distributed participant votes, suggesting unclear emotional cues, while low entropy indicates consistent recognition. This metric allowed the identification of images that were either clearly interpreted or highly ambiguous in conveying the intended emotion. As discussed in Section 4.1, most assignments aligned with ground truth, but two images stood out as outliers. Interestingly, both images ranked among the highest in ambiguity in our user study: Figure 2 (a) was the most ambiguous for emotion selection (Figure 5, b) and scored the 5th highest entropy regarding sentiment. As opposed to those results, Figure 5 (a) shows a low entropy, i.e., low emotion ambiguity image. Figure 2 (b) was among the most ambiguous items, ranking 9th in emotion entropy and 4th in sentiment entropy. Notably, it was the only image in the dataset where participants more often assigned emotions from the opposite sentiment category than the intended one: only 40.9% labeled it as negative, even though it was prompted to convey sadness. Overall, sentiment classification was clearer—likely due to the binary choice provided to participants—though some ambiguity persisted in certain cases, as illustrated in Figure 6 (b). This is the image with the highest sentiment ambiguity, receiving only 54.6% of votes for the intended positive sentiment. As opposed to this, Figure 6 (a) shows a low entropy, i.e., low sentiment ambiguity image.

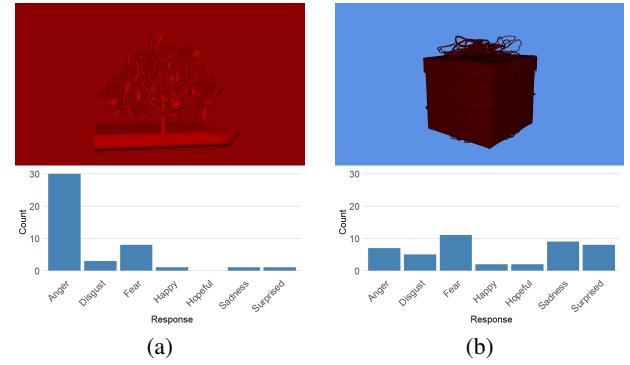


Figure 5: (a) The least (low entropy) and (b) most (high entropy) ambiguous images from the user study for *emotion* selection.

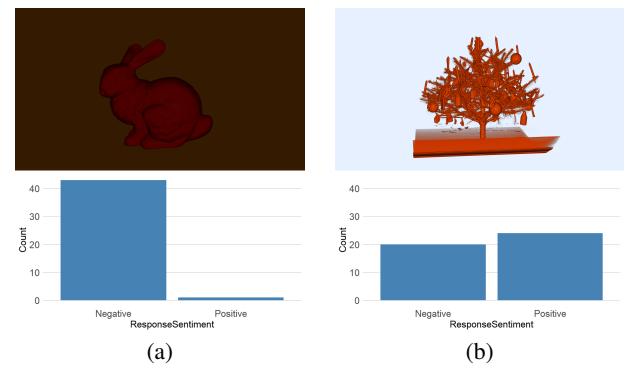


Figure 6: (a) The least (low entropy) and (b) most (high entropy) ambiguous images from the user study for *sentiment* selection.

4.3 Discussion and Conclusions

Our initial results show that LLM-generated visual adjustments are effective in communicating sentiment (negative vs. positive) and can convey specific emotions better than others. Generally,

precise emotion recognition remains challenging. As discussed above, a few ambiguous outlier cases that deviated from the ground truth mappings also revealed structural limits in how reliably LLM-generated parameters convey granular emotions. For instance, we noticed that certain parameter combinations (e.g., subtle variations in brightness) can produce ambiguous interpretations. Entropy analysis on the user study results highlight that emotions like anger and other negative states are relatively consistent and readable, while surprise, fear, and certain depictions of sadness are prone to misinterpretation, often blending with more positive associations. Sentiment (negative vs. positive) association is, in general, more reliable.

Our study has several *limitations* that inform directions for future research. First, our dataset of 28 images and four neutral objects provides only a preliminary basis for assessing LLM-driven emotion rendering. Expanding the variety of objects, contexts, and scene complexities could reveal broader patterns. Second, while we relied on literature-based mappings for validation, emotional associations with color and brightness can vary across cultures, demographics, and individual experience, suggesting the need for more diverse participant pools and cross-cultural evaluation. Third, our approach used a single LLM and fixed prompting. Future work could compare models, prompting strategies, or multimodal approaches to improve robustness and specificity in emotion rendering. Finally, while sentiment recognition was relatively strong, precise emotion recognition remains limited, pointing to opportunities for refinements, such as the use of additional visual cues or adaptive prompting strategies that better capture emotions.

All in all, our work is an initial step toward human-readable emotion generation in 3D visualizations. It highlights both the potential of LLM-guided parameter control and the need for further refinement and evaluation in more complex settings.

REFERENCES

- [1] M. Amencherla and L. R. Varshney. Color-based visual sentiment for social communication. In *2017 15th Canadian Workshop on Information Theory (CWIT)*, pp. 1–5, 2017. doi: 10.1109/CWIT.2017.7994829 1, 2, 4
- [2] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200, 1992. doi: 10.1080/02699939208411068 1
- [3] A. Hanjalic and L.-Q. Xu. Affective video content representation and modeling. *Multimedia, IEEE Transactions on*, 7:143 – 154, 03 2005. doi: 10.1109/TMM.2004.840618 1
- [4] S. Liu, H. Miao, Z. Li, M. Olson, V. Pascucci, and P.-T. Bremer. Ava: towards autonomous visualization agents through visual perception-driven decision-making. In *Computer Graphics Forum*, vol. 43, p. e15093. Wiley Online Library, 2024. 1, 2
- [5] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM International Conference on Multimedia, MM ’10*, p. 83–92. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1873951.1873965 1, 2
- [6] J. A. Mikels, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4):626–630, 2005. 3
- [7] A. Ortis, G. M. Farinella, and S. Battiato. Survey on visual sentiment analysis. *IET Image Processing*, 14(8):1440–1456, 2020. doi: 10.1049/iet-ipr.2019.1270 1
- [8] S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León. Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior Research Methods*, 48(4):1285–1297, 2016. doi: 10.3758/s13428-015-0641-6 1, 2, 4
- [9] K.-C. Peng, T. Chen, A. Sadovnik, and A. C. Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 1, 2
- [10] R. Plutchik. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pp. 3–33. Elsevier, 1980. 1
- [11] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977. 1
- [12] P. Valdez and A. Mehrabian. Effects of color on emotions. *Journal of experimental psychology: General*, 123(4):394, 1994. 1
- [13] J. Z. Wang, S. Zhao, C. Wu, R. B. Adams, M. G. Newman, T. Shafir, and R. Tsachor. Unlocking the emotional world of visual media: An overview of the science, research, and impact of understanding emotion. *Proceedings of the IEEE*, 111(10):1236–1286, 2023. 1
- [14] X. Wang, J. Jia, and L. Cai. Affective image adjustment with a single word. *The Visual Computer*, 29(11):1121–1133, 2013. 1

APPENDIX

The prompt used for the response generation is the following:

In the following messages, you will receive prompts containing emotions as natural language words. It is your task to apply them to a 3D scene containing a single object by adjusting brightness (float between 0 and 1), object color (hex code), and background color (hex code).

Brightness controls the intensity of light on the object. A brightness of 0 results in a completely black silhouette with no visible details or color. A brightness of 1 results in a brightly lit object in its selected color. Values below 0.3 create extremely dim lighting where the object's color becomes difficult to perceive. Since color perception is important to this task, avoid using brightness values below 0.3 for all emotions. The background is unaffected by changes in brightness and will always appear as the color you select. The object is fully visible, carries little emotional association, and is centered in front of a solid color background.

Use the following findings to base your value selections on:

Positive sentiment can be expressed through moderate or higher brightness and high saturation. Negative sentiment is associated with low saturation and extremes of brightness. For low-frequency parts of the scene, in this case, the background, colors relating to cool weather are associated with positive sentiment. Warm tones in the background are more often found in images with negative emotional connotations.

The following color-to-emotion associations resulted from a survey presented in the paper "Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection". The upper 90% of selected colors and their selection frequency are provided; use these values to determine a fitting color for the object:

*Fear - Red 49%, Black 33%, Gray 6%, Blue 5%
Disgust - Green 25%, Brown 21%, Red 20%, Black 18%, Gray 4%, Blue 3%
Sadness - Blue 60%, Black 19%, Red 7%, Gray 4%, Yellow 4%
Anger - Red 79%, Black 14%
Happy - Yellow 53%, Red 17%, Pink 11%, Blue 8%, Green 6%
Surprised - Red 32%, Yellow 26%, White 9%, Blue 8%, Orange 8%, Pink 7%
Hopeful - Blue 29%, White 28%, Yellow 18%, Pink 9%, Red 5%, Orange 3%*

Provide the selected values in this format:

*Brightness: [FLOAT]
Object Color: [HEX]
Background Color: [HEX]*

Legend:

- █ General information on task, parameter behavior, and environment
- █ Findings from Amencherla and Varshney [1]
- █ Findings from Palmer et al. [8]