

# Seeing the Surreal: Mapping Surrealism in Photorealistic AI-Generated Images Using Large Language Models

Xinyi Liu<sup>a</sup>, Yingdan Lu<sup>a</sup>, Qiyao Peng<sup>b</sup>, Sijia Qian<sup>c</sup>, Yilang Peng<sup>d</sup>, Cuihua Shen<sup>e</sup>

<sup>a</sup> Department of Communication Studies, Northwestern University, United States

<sup>b</sup> Department of Communication, University of California, Santa Barbara, United States

<sup>c</sup> School of Data Science, University of North Carolina, Charlotte, United States

<sup>d</sup> Department of Financial Planning, Housing and Consumer Economics, University of Georgia, United States

<sup>e</sup> Department of Communication, University of California, Davis, United States

## Abstract

Photorealistic AI-generated images (AIGIs) are increasingly indistinguishable from real photographs, raising significant social concerns. While prior research focuses on the production quality and detection of photorealistic AIGIs, such research often overlooks their expressive features. This study focuses on surrealism as a key feature of AIGIs, and introduces the concept of algorithmic surrealism to capture AIGIs' algorithmically driven and public accessible generative processes and consequences. Using 28,290 AIGIs collected from Instagram creators and a mixed-methods, Large Language Model (LLM)-assisted framework, we categorized physical, behavioral, and contextual surrealism at scale and found a pervasive presence of surrealism in AIGIs. Topic network and qualitative analyses show that algorithmic surrealism often appears in hybrid forms, indicates patterns of visual excess, reinforces stereotypes, transforms technical flaws into surreal aesthetic features, and exhibits visual homogenization tendencies. This study advances the theoretical understanding of surrealism and photorealism in the age of generative AI. Methodologically, it contributes to computational social science by demonstrating an LLM-based framework that integrates computational, qualitative, and network analyses to examine complex visual concepts.

**Keywords:** AI-generated images, surrealism, photorealism, visual, Large Language Model

## Introduction

Recent advances in text-to-image models like Midjourney and DALL-E have produced photorealistic AI-generated images (AIGIs) that are increasingly indistinguishable from real photos, raising significant concerns. In 2023, AIGIs depicting Donald Trump getting arrested and Pope Francis in a Balenciaga puffer jacket spread widely on social networks, misleading many social media users (Devlin & Cheetham, 2023). In 2025, a White House post depicting Trump as the Pope sparked backlash among Catholic communities (Kavi, 2025). These cases suggest that AIGIs not only blur the line between the real and the artificial but also function as surreal visual artifacts that present objects that deviate from reality in implausible and illogical ways, which attract online attention and shape public perceptions (DiResta & Goldstein, 2024).

Despite growing interest in photorealistic AIGIs, existing research has largely focused either on the technical capability of producing images that are virtually indistinguishable from real photographs, or on examining how humans distinguish synthetic images from real photographs based on visual cues (Z. Chen et al., 2024; DiResta & Goldstein, 2024; Kamali et al., 2024, 2025; Menczer et al., 2023; Q. Peng et al., 2025). However, these studies tend to treat surreal elements primarily as artifacts to be detected or as misinformation cues to be eliminated, overlooking their artistic and communicative potential to attract attention and convey meanings. This paper foregrounds surrealism—a long-established concept in art history and visual communication—as a theoretical framework for understanding the expressive value of AIGIs (Homer & Kahle, 1986; Messaris, 1996). Traditional surrealism is characterized by strategies such as the juxtaposition of unrelated objects and the manipulation of their attributes (Homer & Kahle, 1986). In the generative AI era, however, deep learning and generative models extend the concept of surrealism to a new form—algorithmic surrealism—shaped by algorithmic processes and the participation of both artists and everyday users. In this paper, we aim to conceptualize, identify, and interpret algorithmic surrealism in the context of photorealistic AIGIs.

We constructed a novel dataset of 28,290 AIGIs collected from Instagram AIGI creators, of which 26,771 are photorealistic AIGIs. We developed a mixed-method, large language model (LLM)-assisted framework that can (1) identify and categorize surrealism at scale and (2) examine common visual elements within these surreal AIGIs. Combining human annotation, qualitative analysis, and LLM-based annotation, we identified three categories of surrealism: *physical surrealism*, *behavioral surrealism*, and *contextual sur-*

*realism*, and examined their prevalence, with particular attention to AIGIs featuring political figures. Using LLM-assisted visual summary generation, topic network analysis, and qualitative interpretation, we identified key visual elements across surreal AIGIs, analyzed their co-occurrence patterns, and interpreted how these elements express algorithmic surrealism.

This study offers several key contributions. First, we conceptualize algorithmic surrealism in AIGIs as a contemporary extension of surrealism and delineate its unique features and consequences, addressing the gap in theorizing surrealism in the age of generative AI. Our categorization provides one of the first theoretical frameworks for systematically understanding the visual characteristics of algorithmic surrealism. Our findings empirically demonstrate how algorithmic surrealism extends beyond traditional surrealism rooted in the human unconscious. We also extend theories of photorealism by demonstrating how algorithmic surrealism leverages photorealistic aesthetics to render conceptually implausible content visually credible.

Second, through a large-scale, real-world dataset from Instagram, we provide a comprehensive, content-oriented empirical analysis of AIGIs that shifts the focus from production and detection to the interpretation of visual content. This perspective encourages future research to consider AIGIs not only as objects of verification but also as creative artifacts for interpretation. The emphasis on AIGI content lays the groundwork for future work on perception and deception in generative media.

Methodologically, we introduce a mixed-methods, LLM-assisted framework that integrates quantitative annotation, qualitative interpretation, and topic model network analysis. Our pipeline demonstrates a workflow that moves from human-driven codebook development and iterative prompt engineering to large-scale LLM annotation, combining both top-down, typology-driven annotation and bottom-up summary generation for subsequent unsupervised analysis. This approach provides a scalable and replicable path for analyzing high-level, abstract visual concepts that are difficult to capture using traditional computer vision techniques.

Finally, our work has practical implications for AI literacy and governance. The proposed typology provides media literacy frameworks for future inoculation research and AI literacy programs. Our findings on the surreal depiction of public figures highlight the urgent need for updated ethical guidelines, transparent labeling standards, and effective content moderation protocols—particularly in the context of persuasive or political communication. More broadly, this research carries significant implications

for rethinking the relationship between human and AI creators, raising new questions about the nature of artistic creation, originality, and the tension between human and algorithmic imagination.

## Literature Review

### AIGI and Photorealistic AIGIs

AIGIs are produced by artificial intelligence using algorithms that automate large-scale content production (Cao et al., 2023). Unlike traditional photographs that use a camera to capture real scenes and moments (Hausken, 2024; Lyu & Farid, 2005) or AI-enhanced photos that edit certain parts of a real photograph (Patel, 2024; Williams, 2023), AIGIs are created entirely through algorithmic processes, without any real-life camera captures. While recent research has shown a growing interest in the features and themes of AIGIs on social media (Y. T. Chen & Zou, 2023; Cozzolino et al., 2024; Luccioni et al., 2023; Somepalli et al., 2023), much of this work remains skewed toward artistic or stylized AIGIs, or is limited to specific aspects such as image quality or production-related biases.

In this paper, we examine real-world photorealistic AIGIs that circulated on visual social media platforms. *Photorealism* refers to a visual style in which scenes or objects present a high degree of realism (Griffin, 2025; Joon, 2010). It has been explored across diverse media, such as computer graphics (Ferwerda, 2003; Joon, 2010), artistic works (Griffin, 2025; Hausken, 2024), and mixed and augmented reality (Pereira et al., 2021). While early twentieth-century photorealism in painting emphasized artistic skills by producing paintings that could mimic photographs, the concept later expanded beyond art and was applied to diverse media such as advertising, video games, and computer-generated images (Griffin, 2025). We focus on photorealism in AIGIs, which typically features visual cues such as natural lighting, photorealistic surface textures, and lens artifacts that collectively evoke the look and feel of camera-captured photographs (Hausken, 2024; Wasielewski, 2024).

With recent advances in generative models, such as Google's Flash Image and Imagen, OpenAI's GPT-4o and DALL-E, and Midjourney, photorealistic AIGIs have become increasingly indistinguishable from real photographs, offering high levels of visual detail at significantly lower cost and effort (Devlin & Cheetham, 2023; Hausken, 2024; Wasielewski, 2024). These models far surpass earlier image-generation techniques in their ability to produce highly realistic visuals and are more likely to lead viewers to perceive synthetic

images as real.

As a result, photorealistic AIGIs have drawn significant attention for their role in visual misinformation. Previous studies have documented the spread of viral photorealistic AIGIs on social media platforms (Y. T. Chen & Zou, 2023; Devlin & Cheetham, 2023), showing how they can mislead viewers, erode trust in photographic evidence, and serve as tools for spreading disinformation (Growcoat, 2023; Kamali et al., 2025; Q. Peng et al., 2025).

Research on photorealistic AIGIs generally falls into two main areas of focus. The first centers on the production side, exploring how generative AI models are capable of producing AIGIs that are virtually indistinguishable from real photographs (Z. Chen et al., 2024; Lu et al., 2023). The second line of inquiry investigates user discernment, examining how humans attempt to distinguish AIGIs from real photographs based on visual cues (Kamali et al., 2024, 2025).

We argue that research on photorealistic AIGIs should move beyond capabilities and detection and instead explore their content patterns. In particular, photorealistic AIGIs serve as a visually credible canvas for *surrealism*, a long-standing visual tradition that has become increasingly salient in the context of generative visuals. Unlike illustrations or cartoons—where surrealism and exaggeration are expected—the surreal elements in photorealistic AIGIs are more perceptually striking because they closely mimic real-world scenarios. In this way, photorealistic AIGIs occupy a critical space where realism and surrealism intersect, challenging viewers' assumptions about realism and inviting new forms of interpretation.

## **Algorithmic Surrealism: A New Form of Visual Surrealism in the Age of Generative AI**

Surrealism emerged in the 1920s as a philosophical and aesthetic movement that sought to liberate thought from rational control. Surrealist artists, following how Breton (1969) originally defined surrealism as “psychic automatism in its pure state”, emphasized the spirit of forgoing conscious reasoning, rejecting rational ways of perceiving the world, and the pursuit of inspiration through access to the unconscious mind. Surrealist imagery became one of the most recognizable expressions of this movement, challenging aesthetic conventions and reconfiguring meaning through visual representation (The Museum of Modern Art, 1994). In painting, classic examples include Salvador Dalí's *The Persistence of Memory*, in which melting clocks distort the viewer's sense of time and space. In photography, surrealist artists used various techniques to produce dreamlike effects. A notable

example is Philippe Halsman's *Dalí Atomicus*, one of the most famous surrealist photographs, depicting Salvador Dalí floating in midair alongside flying cats, suspended water, an easel, and other studio objects.

With the advent of computers, surreal imagery has proliferated in digital forms, using computers, graphic tablets, and software tools. Unlike how Philippe Halsman relied on human assistants to hold chairs and toss cats to produce *Dalí Atomicus*, digital artists can achieve similar dreamlike effects more seamlessly through Photoshop and other photo editing software. Advances in computer graphics have further enabled artists to employ 2D and 3D rendering techniques to generate surreal computer-generated imagery (CGI) that is illogical or imaginative. These CGIs have become prevalent across graphic design, video games, and numerous other media forms (A. Smith & Hutson, 2024).

Contemporary advances in deep learning and generative adversarial models have extended the concept of surrealism in the generative AI era, giving rise to a new form of surrealism—*algorithmic surrealism*—a contemporary form shaped by algorithmic and data-driven processes (Halperin & Lukin, 2024; Kratky, 2022; Schröter, 2023). This new form retains core surrealist principles and aesthetics, but fundamentally departs from traditional surrealism in its production mechanisms and downstream consequences.

From the production side, unlike early surrealists who accessed their unconscious space for art creation, surreal AI imagery is generated through algorithmic automatism, that is, the automation of decisions by algorithms. Diffusion networks and transformer-based architectures are trained on massive real-world images and operate stochastically. Rather than “understanding” their outputs, these models predict and recombine visual patterns learned from heterogeneous training data encompassing multiple visual cultures and styles (Croitoru et al., 2023). The generative process itself autonomously combines and composes irrational, dreamlike elements, with humans mostly involved in prompt specification and output selection. At the same time, the ease of producing AIGIs through simple text prompts significantly lowers the barrier to creating surreal imagery, enabling participation not only by professional artists but also by everyday users.

These algorithmically driven, publicly accessible production mechanisms give rise to two downstream consequences of algorithmic surrealism. The first is an expansion of the surrealist visual field. As ordinary users participate alongside trained artists, surrealist aesthetics extend beyond the institutional boundaries of traditional artistic practice. This expansion is also accompanied by algorithmic excess, manifested through overgen-

eration, hyper-detailing, and structural distortions, as models attempt to reconcile contradictory visual cues learned from vast training datasets (N. Smith & Southerton, 2025). This process gives rise to physical anomalies, implausible gestures, and mismatched contexts that exceed the imaginative or irrational boundaries of traditional surrealism grounded in the human unconscious as its creative source.

However, algorithmically driven production also carries risks of bias and visual homogenization. Because algorithmic automatism operates within data ecosystems shaped by circulation, popularity metrics, and existing representational patterns, surreal AI-generated images often replicate dominant visual motifs and reproduce social biases embedded in training data. For example, gender biases present in source datasets may be transmitted to or amplified within generated imagery (L. Sun et al., 2023). Moreover, rather than producing fundamentally new visual logics, algorithmic automatism frequently recombines existing visual tropes through stochastic processes, potentially leading to the homogenization of visual formats and elements in algorithmic surrealism (AIDahoul et al., 2025).

Beyond its artistic origins, surrealism also holds important implications for communication research. Since the 1930s, surrealism also gained momentum in advertising, transforming visual brand communication through surreal content strategies (Homer & Kahle, 1986). Empirical research shows that absurd or incongruous visuals can facilitate message processing by balancing cognitive assimilation and accommodation (Homer & Kahle, 1986), shape brand attitudes (Arias-Bolzmann et al., 2000; Gelbrich et al., 2012), and improve product recall (Gelbrich et al., 2012). Surrealism also plays an important role in political communication. For example, during the Gezi Park protests in Turkey, humorous and absurdist imagery was used to expose the perceived irrationality of authoritarian rule and foster solidarity among diverse protest groups (Korkut et al., 2022).

Building on the concept of algorithmic surrealism embedded in AIGIs and the communicative role of surrealism, this study addresses two key research gaps in understanding surrealism within this emerging visual genre. First, existing research on surrealism in AIGIs often overlooks its artistic and expressive dimensions, instead reducing surreal elements to visual implausibility or technical artifacts. In computer science, for example, surrealist features—such as unrealistic object behaviors or cultural inconsistencies—are typically framed as indicators of generation error, particularly in the context of photorealistic AIGIs (Kamali et al., 2024, 2025).

However, framing algorithmic surrealism solely as a technical defect

obscures its expressive and communicative potential within photorealistic AIGIs. Increasingly, synthetic images are not just produced to mimic reality; they are also optimized for virality, attention, and entertainment (DiResta & Goldstein, 2024; Wei & Tyson, 2024). In other words, algorithmic surrealism operates not as a flaw, but as a content feature that can enhance visibility and engagement. Thus, there is an increasing need to engage surrealism as a meaningful content feature within algorithmically mediated visual communication.

Second, despite its historical significance, few studies have systematically examined content categories or common visual elements of surrealism across different types of visual media, and even fewer have done so for algorithmic surrealism in photorealistic AIGIs. As algorithmic surrealism is distinct from traditional surrealism in both production mechanisms and content, applying prior frameworks developed for other visual formats directly to photorealistic AIGIs may bias analytic results. Although previous studies of AIGIs do not explicitly conceptualize categories of surrealism, some offer promising analytical starting points. For example, Kamali et al. (2024) and Z. Chen et al. (2024) identify several content categories present in AIGIs that align with classic surrealist strategies. For instance, implausible contexts mirror surrealist techniques of dislocation and environmental mismatch; visual appearance exaggerations echo modifications of physical form; and violations of expected functionality parallel the paradoxes and conceptual incongruities described by Z. Chen et al. (2024) and Kamali et al. (2025). Although these taxonomies were originally developed for detection purposes, they can be repurposed to inform a more systematic analysis of the visual features of algorithmic surrealism in photorealistic AIGIs.

Building on these theoretical motivations, we focus on Instagram, a platform where photorealistic AIGIs are rapidly proliferating and driving both viral misinformation and creative production (Devlin & Cheetham, 2023; Growcoot, 2023). In this study, we are interested in what kind of implausibility each image presents, captured through the surrealism categories, and in the semantic and aesthetic patterns that recur across images, referred to as visual elements. Together, these perspectives distinguish between conceptual types of surrealism and their recurring visual manifestations. We propose the following research questions:

*RQ1:* How prevalent is algorithmic surrealism in photorealistic AIGIs from AIGI creators on Instagram, and what are their main categories?

*RQ2:* Besides major categories, what are the common visual

elements of surrealism in photorealistic AIGIs?

## **Applying LLMs in Understanding Algorithmic Surrealism in AIGIs**

Previous research has presented a range of methods for understanding the content of AIGIs. One approach involves quantitative annotation coupled with qualitative taxonomy development, aiming to identify visual features that help human viewers distinguish AIGIs from real photographs (Kamali et al., 2024, 2025; Q. Peng et al., 2025). Some studies have employed computer vision methods to investigate human facial patterns in AIGIs, such as gender biases embedded in image-generative models (L. Sun et al., 2023). Other works have leveraged user perceptual experiments to evaluate how viewers respond to AI-generated visuals (Z. Chen et al., 2024). However, these approaches are often constrained by scalability, manual coding demands, and limited capacity to capture abstract or emergent patterns in generative visuals. They typically rely on predefined codebooks and lack the flexibility to account for stylistic novelty or conceptual ambiguity (Joo & Steinert-Threlkeld, 2022).

To address these limitations, our study introduces an LLM-supported approach that combines LLM approaches with conventional qualitative and quantitative methods. Recent advances in LLMs demonstrate their potential for scalable content analysis, as they increasingly align with human judgment in interpreting textual data (Wu et al., 2023). In computational social science, LLMs have been widely adopted for tasks such as text extraction (Stuhler et al., 2025), document annotation (Chew et al., 2023; Rathje et al., 2024), and summarization (Pham et al., 2024), offering both efficiency and interpretability across large corpora.

In addition to textual data, LLMs also show powerful capabilities in understanding visual messages. Multimodal LLMs, such as GPT-4o and Gemini, have been increasingly applied in visual content analysis. Recent computer science research has demonstrated LLMs' strong visual capabilities in understanding, detecting, and extracting concepts and features about objects or interactions in visual or multimodal forms (Cui et al., 2023; Liu et al., 2023; Ma et al., 2024; Schulze Buschoff et al., 2025). Furthermore, some benchmark studies also show how state-of-the-art LLMs are able to handle abstract and ambiguous visuals that require stylistic background and cultural knowledge (Shahgir et al., 2024; Xiao et al., 2025). In social science, emerging works have begun leveraging multimodal LLMs to understand real-world images, such as social media visuals and political video content (Breuer et al., 2025;

Y. Peng et al., 2025).

In addition, the generative and reasoning capabilities of LLMs distinguish them from conventional visual analytics methods. Unlike traditional approaches that primarily output numerical results, LLMs can produce detailed, natural language explanations of how they interpret images, offering greater interpretive transparency and flexibility (Chew et al., 2023). These outputs can also be repurposed for downstream analyses.

Prior applications of LLMs in visual studies generally follow two approaches. The first is a supervised approach, which prompts LLMs with direct visual inputs alongside predefined coding schemes, enabling a top-down feature extraction and analysis through zero-shot or few-shot settings (Y. Peng et al., 2025; Schulze Buschoff et al., 2025). In this context, LLMs not only automate annotation but also assist in refining codebooks during the development phase by providing concept-level suggestions and justifications (Chew et al., 2023).

The second is an unsupervised approach, in which LLMs generate textual summaries directly from visual inputs that can then be analyzed using natural language processing (NLP) tools such as topic modeling (Y. Peng et al., 2025; Y. Sun et al., 2025). This bottom-up approach enables feature discovery from large-scale visuals. For instance, Y. Sun et al. (2025) employed LLMs to generate descriptions of internet memes and used these summaries as input for topic modeling. This approach leverages LLMs' ability to translate visual content into semantically rich text, making it easier to identify latent themes across large image corpora. By converting visuals into language, LLMs facilitate a form of analysis that is both scalable and interpretable, while also enabling the discovery of emergent patterns without relying on predefined coding categories. Moreover, it supports iterative, human-in-the-loop coding processes that improve interpretive robustness over time (Y. Peng et al., 2025).

Acknowledging these advancements and different pathways in using LLMs to analyze visual concepts, we ask the methodological question:

*RQ3: How can we use large language models to analyze surrealism through both top-down and bottom-up approaches?*

## Data and Methods

### Social Media Data: Photorealistic AIGIs from Instagram

We gathered data from Instagram through a two-step procedure, as Figure 1 shows. First, we identified AIGI accounts that frequently posted AIGIs from

July 2022 to August 2023. We started by looking through news articles that mentioned viral photorealistic AIGIs to identify seed accounts on Instagram where these AIGIs originated. Then, we snowballed and manually collected more AIGI accounts that were mentioned, forwarded to, or appeared on the recommendation lists of those seed accounts. We included AIGI accounts that were either self-identified as such or had at least 50% of their most recent 20 images accompanied by captions, hashtags, or posts indicating that the images were created using specific generative tools. We terminated the sampling process once additional accounts yielded only repeated mentions or recommendations. Two Instagram accounts were excluded due to account deletion, resulting in a final account sample of 47 Instagram accounts. A detailed description of the seed accounts, news reports in which we identified the accounts, and the account snowballing procedures can be found in the Appendix A.<sup>1</sup>

From these accounts, we collected images from their Instagram timeline posts between July 12, 2022 and August 31, 2023. The start date was chosen to align with the public release of the most widely used AI image-generating tools, Midjourney and DALL-E. The end date was set for consistency, since image collection began in September 2023, and we aimed to standardize the end time across all accounts. We used 4K Stogram<sup>2</sup> to download the timeline posts and the associated images, resulting in a total of 28,290 images (Full dataset).

To refine the dataset to include only photorealistic AIGIs, one author manually reviewed all collected images and excluded 1,519 images that were truncated files, non-AIGIs, or non-photorealistic AIGIs. Non-AIGIs included items such as screenshots of social media or software interfaces, as well as text-only images without meaningful visual content. Non-photorealistic AIGIs were defined as images displaying characteristics of artistic illustration, graphic design, or digital painting. This cleaning process yielded a final set of 26,771 photorealistic AIGIs (Sample A) for analysis.

We applied an LLM-assisted framework to identify, categorize, and contextualize surrealism at scale, which comprises two components: (1) a top-down analysis combining human annotation, qualitative coding, and GPT-4o-supported annotation to classify surrealism based on a human-defined typology; (2) a bottom-up Topic Model Network approach that used LLM-generated visual descriptions to uncover broader visual elements in which

---

<sup>1</sup>The Appendix and replication materials are available at OSF: [https://osf.io/mfwbr/overview?view\\_only=6ccb52f4c2864bb7a33eb67d25dfee05](https://osf.io/mfwbr/overview?view_only=6ccb52f4c2864bb7a33eb67d25dfee05)

<sup>2</sup>4K Stogram is a desktop application for downloading photos, videos, and stories from Instagram. <https://www.4kdownload.com/products/stogram-8>

surrealism appears.

## LLM-assisted Annotation of Surrealism

### Defining Surrealism Categories in Photorealistic AIGIs

We started by developing a taxonomy of surrealism in AIGIs through qualitative visual analysis and quantitative content analysis (step 1 in Figure 1). We first sampled 200 AIGIs (Sample B) from the full sample, and designed a simple binary coding scheme called “the presence of surrealism combination,” which captures the presence of unnatural pairings, such as implausible combinations of humans, objects, behaviors, or situations that are not typically observed together in real life, drawing on prior studies (Homer & Kahle, 1986; Kamali et al., 2024, 2025). Based on feedback from this training phase, we refined the coding guidelines with additional examples and clarifications. Two trained coders independently annotated the 200 images, achieving satisfactory intercoder reliability (Cohen’s Kappa = .79). Using the refined coding scheme, we further applied the guidelines to an additional 1,800 AIGIs. Among the 2,000 human-annotated images, 1,795 were photorealistic AIGIs, and 1,254 (69.9%) of the photorealistic AIGIs contained surreal elements.

During coding, we frequently observed recurring elements such as public figures and unusual outfits. This motivated us to develop a systematic categorization that moves beyond detection toward interpretation, and to develop the concept of algorithmic surrealism in AIGIs. We started from a qualitative analysis of all annotated surreal images depicting humans. Coders recorded observations of surreal elements, and the research team met to discuss recurring visual patterns. We then expanded the coding scheme to include non-human subjects (e.g., animals, objects, nature) and refined our categories.

From this process, three distinct forms of surrealism emerged: physical, behavioral, and contextual surrealism.

*Physical surrealism* refers to unnatural or implausible physical characteristics, such as facial abnormalities, distorted anatomy, or hybridized forms composed of unrealistic materials or textures. The upper row of Figure 2 illustrates these instances in our dataset, depicting distorted faces, incongruent clothing, and metallic skin.

*Behavioral surrealism* captures impossible or supernatural behaviors that violate physical laws, biological constraints, or sociocultural expectations. Examples include animals performing human tasks and people

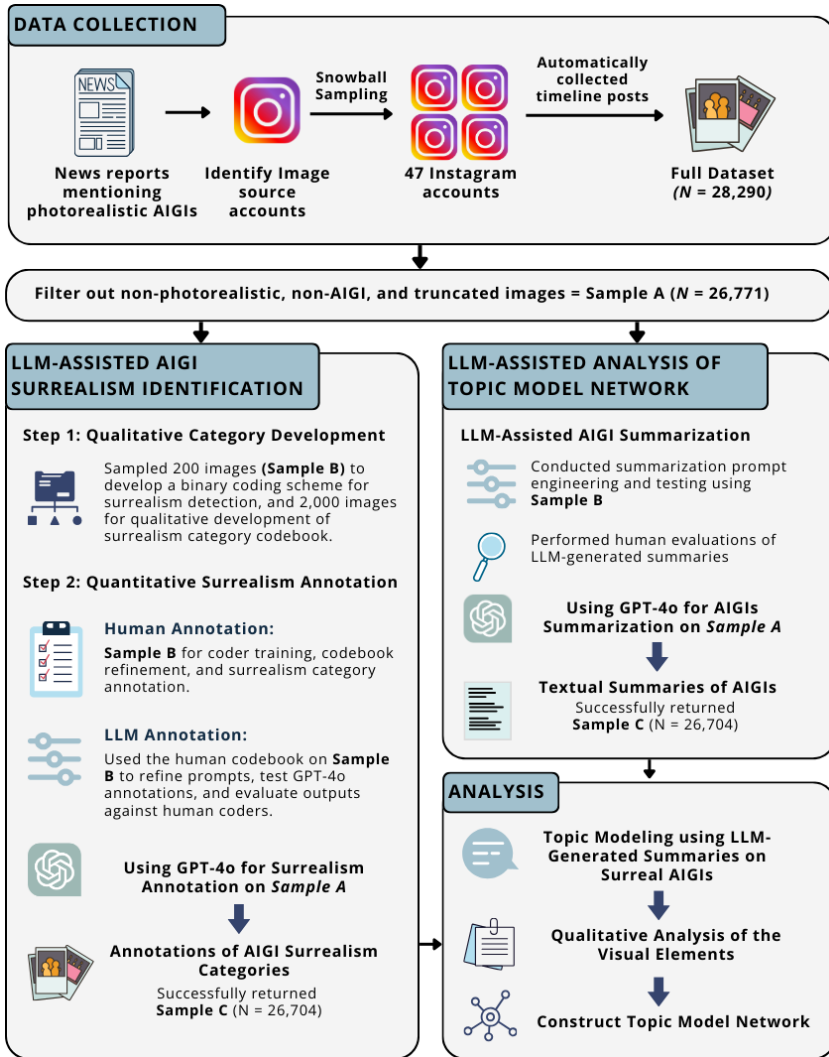


Figure 1: Data Collection and Analytical Methods

floating midair (middle row, Figure 2).

*Contextual surrealism* describes implausible or fantastical settings that violate expectations of time, space, narrative coherence, or plausibility. These scenes often include supernatural landscapes, anachronistic elements, or mismatched spatial dimensions. The bottom row of Figure 2 illustrates this category: an elderly woman in traditional dress with a joyful expression beside a fierce dragon, and Elon Musk standing confidently in close proximity to flames and destruction.

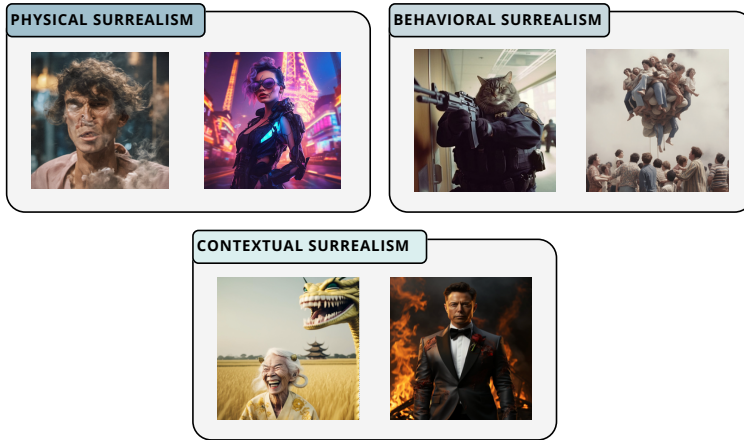


Figure 2: Examples of surrealism in AIGIs

After defining the surrealism categories, we developed a coding scheme to ensure that each category was defined systematically and was replicable (see Table 1). We trained two human coders using the initial codebook, and discrepancies identified during the first-round discussion informed further revisions. Using the finalized codebook, the coders annotated 200 sampled images (Sample B), and all three variables achieved Cohen's  $\kappa$  values above 0.8 (human annotation in step 2 in Figure 1).

### LLM-supported Surrealism Annotation

With the human-coded surrealism categories established, we conducted a computational analysis to examine both the prevalence and characteristics of surreal content across the final set of photorealistic AIGIs ( $N = 26,771$ ). This analysis employed a structured, multi-step annotation pipeline using GPT-4o (model version: gpt-4o-2024-11-20) to annotate each image according to the three predefined surrealism categories through zero-shot, prompt-based

Variable	Codebook Description
Physical Surrealism	Does this image contain any surreal physical characteristics, such as facial or body abnormalities, or unusual clothing? Code 1 for “Yes” and 0 for “No.”
Behavioral Surrealism	Does this image contain any surreal behaviors or interactions, such as supernatural actions, implausible relationships, or unrealistic activities? Code 1 for “Yes” and 0 for “No.”
Contextual Surrealism	Does this image contain any surreal contexts, such as impossible environments, fantasy-like elements, or distorted group depictions? Code 1 for “Yes” and 0 for “No.”

Table 1: Codebook for Surrealism Categories Annotation

modeling.<sup>3</sup>

Building directly on the human annotation codebook, we adapted the same variable definitions to design prompts that instructed GPT-4o to classify the presence of surrealism and its subcategories as binary variables (1 = Yes, 0 = No). Each annotation was accompanied by a brief textual justification, and we also included an additional variable to capture surreal elements not covered by predefined categories. The prompts were finalized through an iterative prompt-engineering process that combined qualitative refinement with comparative evaluation of GPT-4o outputs against human annotations across three evaluation rounds on the same sample coded by human coders (LLM annotation in step 2 in Figure 1). Details of the prompt development, testing, and validation are provided in Appendix B.

Early versions of the prompt revealed systematic errors; for example, GPT-4o tended to over-classify contextual surrealism by interpreting normatively stylized backgrounds as surreal, or under-classify it when the prompt was overly restrictive. To address these issues, we clarified instructions to encourage more careful examination and simplified illustrative examples without altering the variable definition. Through three iterations, the finalized prompt achieved strong alignment with human judgments in detecting

<sup>3</sup>While fine-tuning a pre-trained vision model based on the manually annotated sample could potentially improve classification accuracy, we chose to use a zero-shot prompting approach for three main reasons. First, zero-shot prompting allows the model to remain less task-specific and more generalizable to unseen categories. Second, prompt-based modeling is substantially more cost-efficient than fine-tuning large vision-language models. Third, our methodological goal for this paper is to assess the feasibility of using LLMs to annotate complex visual concepts directly through prompting. This approach allows us to evaluate how well LLMs can perform as zero-shot annotators, contributing new empirical evidence to the growing literature on their utility in computational social science research.

the presence of surrealism. Performance varied slightly by subcategory: Physical Surrealism (Precision = 0.91, Recall = 0.76, F1 = 0.83); Behavioral Surrealism (Precision = 0.73, Recall = 0.73, F1 = 0.73); and Contextual Surrealism (Precision = 0.83, Recall = 0.72, F1 = 0.77). Additionally, no new patterns of surrealism emerged beyond these three categories.

However, due to model constraints, GPT-4o was unable to reliably process AIGIs featuring political figures. In such cases, the model either refused to annotate or failed to detect surreal elements in these images. For example, an image of Joe Biden riding a motorcycle was not correctly identified as surreal because the model interpreted Joe Biden as an ordinary individual rather than recognizing his political identity. To ensure completeness and accuracy, we manually annotated all 467 AIGIs featuring politicians using our predefined codebook ( $N = 467$ ). One primary coder annotated all the politician AIGIs, and a second coder independently verified all labels, and any discrepancies were resolved through discussion to finalize the results. These annotations were used in the analysis of surrealism categories.<sup>4</sup>

## **LLM-assisted Analysis of Topic Model Network for Surreal Visual Elements**

### **LLM-assisted Summary of AIGIs**

To explore the common visual elements used in surreal AIGIs, we instructed GPT-4o to generate concise text descriptions of all AIGIs alongside its annotations for surrealism categories. To evaluate the quality of these LLM-generated summaries, two authors first reviewed 15 randomly sampled images from Sample B alongside their LLM-generated summaries and discussed the resulting human-LLM discrepancies. Based on this discussion, we finalized a codebook with three binary variables capturing potential quality issues in the summaries: missing information, hallucinations, and errors. Missing information refers to key visible elements that were omitted from the summary; hallucinations refer to details not actually present in the image; and errors denote inaccurate descriptions of visible content. Each issue type was assessed along three dimensions of image representation: subject (who/what), action (how), and context (when/where). When a problem was identified, coders recorded brief notes describing the specific inaccuracy

---

<sup>4</sup>While these images were problematic for LLM-based surrealism annotation, we did not encounter similar issues when prompting GPT-4o to generate image descriptions. Therefore, we kept the GPT-generated summaries of these politician images for topic modeling and network analysis.

or omission. Using this codebook, the coders independently evaluated an additional 60 image summaries for the final assessment.

Summary evaluation results show that the quality of LLM-generated summaries was satisfactory in general. The most frequent issue involved missing subjects (17%), followed by subject-related errors (11%), while all other issue types were rare (1–3%). No hallucinations were observed. Appendix C provides illustrative examples of the LLM-generated summaries, corresponding images, and coders' notes.

Together with the surrealism annotation, to maximize data coverage and annotation quality, we ran the GPT-4o annotation in two sequential batches. In the first batch, GPT-4o was allowed up to 20 processing attempts per image, resulting in 4,492 unsuccessful cases. These were processed again in a second batch under the same conditions, reducing the number of failed annotations to 67. After excluding these 67 cases, the final GPT-annotated dataset comprised 26,704 images (Sample C). A detailed pipeline including all data exclusion decisions is presented in Figure A1 in Appendix A.

### **Analysis of Topic Model Network for Visual Elements and Relationships**

To investigate common visual elements embedded in surreal AIGIs, we conducted Analysis of Topic Model Networks (ANTMN) (Walter & Ophir, 2019) based on the textual summaries generated by the LLM for the 17,852 images identified as containing surrealism. This approach integrates the merits of topic modeling and network analysis to identify the prevalent topics, which are recurring visual elements in our context, and examine their co-occurrence across AIGIs to highlight how visual elements are semantically connected in the corpus.

We started by preprocessing the textual summaries, by removing stop-words, converting text to lowercase, removing punctuations and numbers, and excluding words appearing in more than 95% of the documents or fewer than 0.5% (Walter & Ophir, 2019). We also manually removed five words that are overly frequent but not informative in our context (image, depicts, background, person, shows).

For topic modeling, we applied Latent Dirichlet Allocation (LDA) with Gibbs Sampling (Blei et al., 2003; Walter & Ophir, 2019), using the *topicmodels* and *LDA tuning* packages in R for model estimation and evaluation. To identify the optimal number of topics ( $K$ ), we trained LDA models with varying topic counts and evaluated them through 5-fold cross-validation based on held-out perplexity. Candidate models were tested with  $K$  values ranging from 5 to 50 (in increments of 5) and from 20 to 25 (in increments of

2). Through the elbow method, we found  $K = 22$  as the optimal point where additional model complexity provided minimal improvement, representing the best balance between interpretability and performance. All 22 visual elements were manually labeled by one author based on their top 10 representative words and top 10 representative images of each visual element (Roberts et al., 2014). Additional details on the perplexity score, topic number selection, topic modeling process, and the top words are provided in Appendix D.

Then, we constructed the topic model network through four main steps. First, we quantified the relationships among visual elements by calculating the Spearman correlations based on their co-occurrence patterns in the document–topic probability matrix ( $\theta$ ), a standard approach for identifying semantically related topics based on shared textual contexts (Maier et al., 2018). The resulting fully connected, weighted graph was further processed to retain only positive associations (all negative correlations were set to zero) for compatibility with subsequent computation. To reduce noise and highlight statistically meaningful connections, we then applied the Disparity Filter backbone method (Serrano et al., 2009), which retains only edges statistically significant within their local weight distributions. This procedure reduced the network from 163 edges to 32 edges (80.4% removed) while retaining all 22 nodes, yielding a sparse network backbone capturing the most structurally important co-occurrence relationships among visual elements. Third, the topic model network was visualized using Gephi (Bastian et al., 2009), where node size represents the visual element’s overall prevalence (its total probability mass in the corpus) and edge weights indicate visual element similarity. Finally, to characterize the structural roles of different visual elements, we also computed centrality measures on the weighted backbone network, including strength and betweenness centrality. More information on topic model network construction is detailed in Appendix E.

### **Qualitative Interpretation**

Besides quantitative and network analysis, one author conducted a qualitative analysis on all the visual elements, and discussed them among the authors. By reviewing the top-ranked GPT summaries and their corresponding images, this multimodal interpretive step connects the quantitative topic modeling results back to the original data, enabling a richer understanding of how algorithmic surrealism is represented through these visual elements.

## Results

### Prevalence of Surrealism Categories in Photorealistic AIGIs

As Figure 3 shows, surrealism appears in 66.9% of our final sample of 26,704 photorealistic AIGIs ( $N = 17,852$ ), while 33.1% ( $N = 8,852$ ) contain no surreal elements. Among surreal images, physical surrealism is the most prevalent category ( $N = 13,644$ ; 76.4%), followed by contextual surrealism ( $N = 6,279$ ; 35.1%) and behavioral surrealism ( $N = 3,985$ ; 22.3%). Overlaps indicate that surreal categories frequently co-occur: 2,941 images combine physical and contextual surrealism, such as depictions of pedestrians with unusually bent legs walking through tunnels lined with oversized, glowing organic forms, or giant Lego seahorses appearing inside artificial aquariums. 2,602 combine physical and behavioral surrealism, for example, a giant hand extruding a gravity-defying object. 1,498 combine behavioral and contextual surrealism, such as people moving through a cityscape filled with massive floating red cubes. 985 images (5.5% of surreal AIGIs) exhibit all three forms simultaneously. These patterns suggest that physical surrealism constitutes the most common and structurally central category of AI-generated surrealism, while behavioral and contextual elements appear more selectively, often accompanying or amplifying physical distortions rather than occurring independently. Figure 4 provides examples of the images that contain more than one type of surrealism.

Among 467 photorealistic AIGIs featuring political figures, surrealism is nearly universal (98.6%). As with the general dataset, physical surrealism remains most prevalent (79.2%), involving distorted facial features, tattoos, or inappropriate attire that transforms public figures into parodic or humorous representations. For example, as Figure 5 shows, Barack Obama is dressed in a tribal outfit. Contextual surrealism, present in 67.2% of political AIGIs, often places political figures in settings where they would not typically appear, such as fast-food kitchens, underground fighting arenas, public protests, and entertainment venues, such as Donald Trump posing for a group photo in a sci-fi cabin setting (Figure 5). Behavioral surrealism, found in 45.4% of political AIGIs, depicts politicians in physically impossible or socially unexpected actions, such as Joe Biden riding a motorcycle (Figure 5).

### Common Visual Elements in Algorithmic Surrealism

Among the 22 identified visual elements, the most prominent visual element is *Dreamlike and Foggy Scenes* (833 images), characterized by soft lighting,

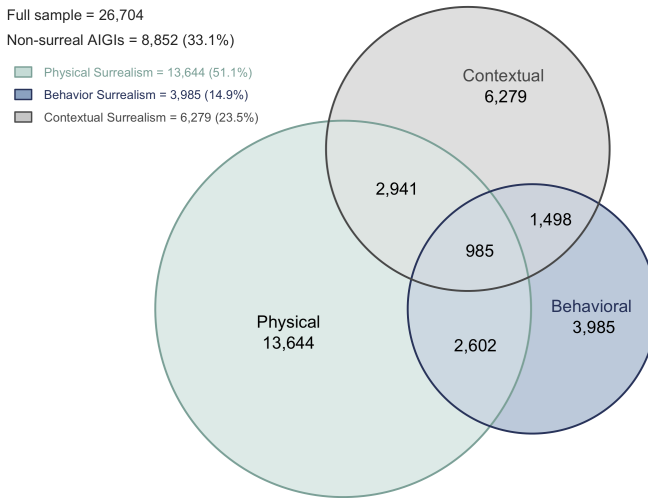


Figure 3: Prevalence and Co-occurrence of Physical, Behavioral, and Contextual Surrealism in AIGIs



Figure 4: Co-occurrence Examples of Physical, Behavioral, and Contextual Surrealism in AIGIs

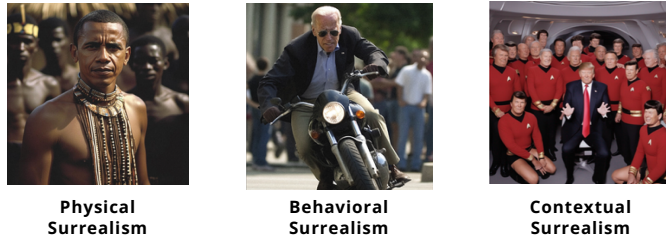


Figure 5: Examples of Surrealism in Photorealistic AIGIs Featuring Politicians

mist, and ethereal illumination that portray a calm and mysterious atmosphere. Another prevalent element, *Unusual Planet Terrains* (818 images), usually depicts barren deserts, rocky mountains, and cosmic landscapes under dramatic skies implying extraterrestrial environments. Beyond these scenic and contextual elements, two other dominant visual elements are *Sci-fi and Futuristic Costumes* (828 images), featuring metallic or high-tech attire frequently situated in cyberpunk or futuristic settings; and *Figures with Masks and Headpieces* (818 images), portraying individuals in elaborate masks, helmets, or full-body suits that obscure identity. Visual examples of these dominant visual elements, as well as a complete list of all visual elements and descriptions, can be found in Appendix D. Figure 6 displays two examples for these top visual elements.

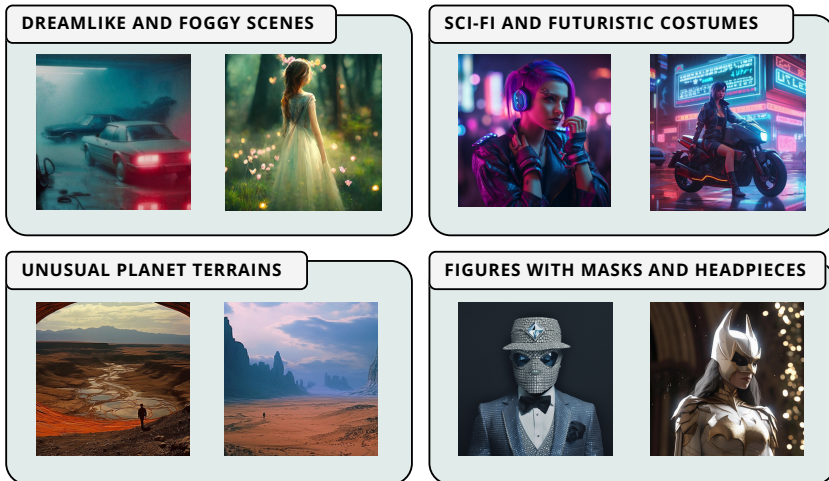


Figure 6: AIGIs Examples in the Most Prominent Visual Elements

Our qualitative analysis reveals that surreal AIGIs exhibit distinct charac-

teristics of algorithmic surrealism in several ways. In *Artwork with Vibrant Colors*, the top images exhibit extremely high levels of colorfulness, contrast, and saturation, while in *Overdecorative Fashion*, the images feature exaggerated textures, intricate fabrics, and ornate embellishments. Together, these patterns create excessive visual complexity, exemplifying the algorithmic excess produced by the model's algorithmic automatism.

Another important qualitative finding is that several visual elements reproduce or subtly reinforce existing social and gender stereotypes. In *Sci-fi and Futuristic Costumes*, nearly all representative images feature female figures depicted in tight, metallic, or body-conforming attire, reinforcing the objectification of women and the gendering of futurism. Similarly, *Cyborgs and Humanoids* frequently portray female bodies as sites of technological transformation. *Artistic Makeup and Face Design* further extends this logic from bodies to faces, treating female faces as canvases for visual experimentation and reinforcing female objectification under the guise of creativity. This pattern reflects the dynamics of large-scale data training and algorithmic automatism in AIGIs, where popular visual themes are more likely to be reproduced, contrasting with traditional surrealism's aim to challenge rationality and social norms.

Finally, our qualitative analysis shows how algorithmic flaws evolve into surreal visual elements, going beyond the imaginative boundaries of traditional surrealism. In *Body and Facial Distortion* and *Crowded Public Gatherings*, many images feature distorted, duplicated, or partially merged human figures, creating an uncanny sense of overpopulation and confusion. These uncanny qualities, such as visual distortions, anatomical inconsistencies, or generation flaws, transform technical limitations into artistic features, making imperfection itself a surreal aesthetic. These findings further indicate that the imaginative scope of algorithmic surrealism extends beyond traditional surrealism grounded in the human unconscious.

## Inter-correlated Surreal Visual Elements

As Figure 7 shows, the network of visual elements is densely connected, indicating that surrealist features often co-occur across AIGIs rather than appearing in isolation. Strong connections emerge between visual elements depicting human figures and those representing environmental settings. For instance, *Sci-fi and Futuristic Costumes* is closely linked with *Unnatural Scene-Object Combination* and *Cyborgs and Humanoids*, suggesting an integrated portrayal of costumed figures within unnatural or artificial settings. Likewise, *Unusual Planet Terrains* is closely connected to *Portraits*

with *Stylized Outfits* and *Body and Facial Distortion*, showing stylized or distorted human forms situated in extraterrestrial environments. We also observe strong ties between *Geometric Shapes and Color Blocks* and *Miniature Worlds*, implying that even simple geometric elements are situated within imaginative contexts.

Among all visual elements, *Cyborgs and Humanoids* emerges as the most structurally central node in the network. It exhibits the highest strength (weighted degree) as well as the highest betweenness centrality, indicating both strong overall connectivity and a key bridging position across thematic clusters. As reflected visually in Figure 7, this element connects human-centered motifs (e.g., *Portraits with Stylized Outfits*) with nonhuman, monstrous, and scenic elements (e.g., *Fantastical Monsters, Surreal Interiors and Rooms*), effectively linking bodily, spatial, and stylistic dimensions of algorithmic surrealism. Beyond this primary hub, *Portraits with Stylized Outfits* and *Unnatural Scene-Object Combination* also occupy structurally central positions, reinforcing the role of stylized human figures and scene-level incoherence as organizing principles of surreal AIGIs. Together, these patterns suggest that algorithmic surrealism is not composed of isolated visual tropes, but rather emerges from hybrid configurations that merge bodies, environments, and artificial aesthetics, giving rise to a transgressive futuristic visual logic.

These visual element co-occurrence patterns also reflect a visual homogenization tendency. The most central nodes in the network, such as *Cyborgs and Humanoids* and *Dreamlike and Foggy Scenes*, show that many surreal AI-generated images converge around and repeat existing visual tropes. These patterns suggest that algorithmic surrealism is likely to be produced by recurrent configurations that merge bodies, environments, and artificial aesthetics. Rather than generating fundamentally new visual elements or forms, many AI-generated images reproduce and recombine established visual elements and formats to construct surreal imagery.

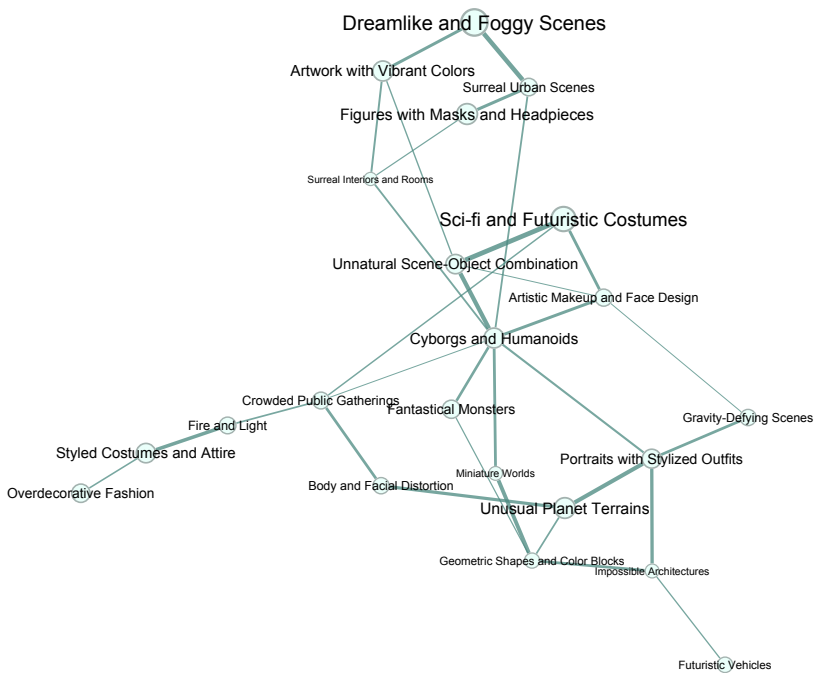


Figure 7: **Topic Model Network Visualization**

*Note.* Node size represents topic prevalence, defined as the mean document-level topic proportion across the corpus; edge weights represent pairwise similarity between topics, calculated as Spearman correlations between document-level topic prevalence vectors. The topic similarity network is a weighted, undirected network.

## Discussion

Grounded in the history and theory of surrealism and its visual manifestations, this study examines surrealism in photorealistic AIGs, and we argue that surrealism in AIGs has evolved into a new form—algorithmic surrealism. Using a mixed-methods, LLM-assisted analysis of 28,290 AIGs collected from Instagram AI creators, we find that surrealism is not a marginal phenomenon but a defining aesthetic feature: approximately 67% of the photorealistic AIGs contain surreal elements. We identify three major categories—physical, behavioral, and contextual surrealism—with physical surrealism being the most dominant, and these categories frequently co-occurring within individual images. Our topic modeling analysis, complemented by qualitative examination of visual elements, reveals that surreal AIGs exhibit algorithmic automatism through excessive detail, visual exaggeration, and implausible bodily or spatial configurations. The topic network reveals that visual elements are densely interconnected in shaping surrealism, and the central elements reproduce established visual configurations and elements.

This study advances the theoretical understanding of surrealism in the age of generative AI. While surrealism has long been examined in traditional arts and early computer-generated imagery, our work highlights the emergence of algorithmic surrealism in visual communication, a form shaped by algorithmic automatism rather than by direct access to the human unconscious. Our findings illustrate the algorithmically driven, publicly accessible surrealism production mechanisms and their downstream consequences. First, the transformation of algorithmic flaws into artistic features demonstrates the stochastic production process and how algorithmic surrealism expands the imaginative scope of the surreal beyond the human unconscious, toward the boundaries of computational systems themselves. Second, our findings on the depiction of politicians, the reproduction of gender stereotypes, and the tendency toward visual homogenization in surreal AIGs demonstrate how algorithmic surrealism reflects the dynamics of large-scale data training, often reproducing dominant visual themes and embedded social biases. In this sense, algorithmic surrealism may expand artistic participation and scale while simultaneously functioning as a form of mass-produced recombination of earlier surrealist tropes. Together, these findings underscore the need for future research on how these visual constructions were built and the underlying mechanisms of AIGI creation.

Through a large-scale, real-world dataset from Instagram, we provide a comprehensive, content-oriented empirical analysis of AIGs that shifts the focus beyond production and detection to the visual content itself. By

categorizing and analyzing surrealism as a content feature of AIGs, we reposition surrealism not just as a technical artifact that is crucial in discernment, but as a communicative medium with distinctive content dimensions. This shift invites scholars to treat AIGs not only as problems to be detected, but as creative artifacts to be interpreted.

From the literature on photorealism, our findings extend existing understandings by showing how surrealism can be embedded within photorealistic aesthetics. By identifying physical, behavioral, and contextual forms of surrealism, we demonstrate that AIGs often mimic real-world scenarios through visually plausible yet conceptually implausible combinations. This reveals an inherent tension between photorealism and surrealism: rather than undermining photorealism, surreal elements exploit it by making the implausible appear believable, such as the impossible juxtaposition of Barack Obama wearing a traditional tribal outfit. Surreal elements may render the implausible believable by drawing on the inherent visual authority and credibility of photorealistic imagery, as viewers may tend to perceive images that conform to photographic conventions as real. In this sense, photorealism in the age of AI becomes not only a visual feature but also a container for surrealism, enabling the new algorithmically uncanny, where photorealism and implausibility coexist within the same image.

Methodologically, this study introduces a mixed-methods, LLM-assisted framework that integrates computational analysis, qualitative analysis, and network analysis to examine surrealism in AIGs, offering key methodological insights for analyzing visual content at scale, especially beneficial for abstract visual features like surrealism. First, by establishing clear and replicable subcategories of AIGI surrealism, this study provides a methodological framework for future research on surrealism in AIGs and related abstract visual phenomena.

Second, we demonstrate the value of a human-involved, LLM-assisted content analysis pipeline: beginning with qualitative insights and human annotation, proceeding to quantitative codebook construction, prompt engineering, and testing, and extending those insights across a large dataset using GPT-4o for structured annotation and summarization. This workflow balances interpretive depth with scalability.

Finally, our work contributes to the growing literature on using LLMs for visual content analysis (Y. Peng & Lu, 2023; Y. Peng et al., 2025; Y. Sun et al., 2025). Unlike traditional computer vision approaches that may struggle with complicated visual concepts like surrealism, LLMs enable higher-level semantic understanding and reasoning across visual and textual formats.

In our case, GPT-4o successfully identified surreal categories, generated effective summaries, and enabled downstream topic modeling and network analysis, offering a more flexible and interpretable approach compared to conventional methods. Future researchers can adopt a similar framework that leverages LLMs for both top-down supervised typology-based annotation and bottom-up unsupervised summary generation to explore similar questions across other image genres or contexts.

Surrealism is not just a theoretical curiosity; it has practical significance in different communication fields and contexts. In visual communication, AIGIs with surreal elements featured in political campaigns or persuasive content are reshaping how audiences perceive visual credibility. The prevalence of human figures in our dataset, especially politicians rendered with a highly professional and photorealistic aesthetic, underscores the potential risks of such imagery being misused for visual misinformation. For instance, the surreal depiction of political figures can blur the boundary between authenticity and fabrication, potentially distorting public perceptions and eroding trust.

The prevalence of algorithmic surrealism in photorealistic AIGIs and their potential risks as well, as algorithmic bias, in visual communication underscore the need for media and AI literacy frameworks that go beyond factual accuracy or tampering detection. Our theorization of algorithmic surrealism across production mechanisms and downstream consequences, together with our empirically grounded typology, contributes to public understanding by identifying the major categories and visual strategies of surrealism in AIGIs. It also provides a foundation for future inoculation research and AI literacy programs.

The occurrence of algorithmic surrealism and its visual manifestations also carries significant implications for rethinking the relationship between human and AI creators. While generative models are trained on massive datasets of real-world images, this process grants greater agency to machines in composing surrealist elements, while the human role increasingly centers on prompt design and curation. This shift reconfigures the creative division of labor between humans and machines and raises new questions about the nature of artistic creation, originality, and the tension between human and algorithmic imagination.

In addition, there is also a critical need to closely monitor and regulate AI-generated content (AIGC) for more ethical use. Building on our typology of surrealism, we propose that ethical guidelines for AIGC and AIGIs should account for different surrealism categories and their communicative effects.

Multiple efforts can be made to improve ethical use, for example, clearer labeling standards, disclosure requirements, and moderation protocols for AI-generated visuals, particularly those featuring algorithmic surrealism that are more likely to be leveraged for visual misinformation. At the same time, AI and social media companies should implement strict rules that govern the inclusion of a real person's images in AI-generated content, ensuring that such usage does not contribute to the dissemination of false information or harm the reputations of those depicted.

The study has several limitations. First, our dataset focuses on Instagram accounts that self-identify as AIGI accounts, which may overrepresent certain genres and underrepresent others. Second, although GPT-4o performed well overall in analyzing AIGIs, it encountered difficulties when processing politically sensitive content, often requiring manual intervention. In particular, our manual re-annotation of political figures after 40 failed trials highlights the limitations of relying solely on automated methods for analyzing sensitive or high-stakes visual content. Future research should explore open-source models to mitigate this issue. Third, given the well-documented social and political biases in LLMs (Bandaru et al., 2025; Farrell et al., 2025; Lee et al., 2024), our analysis of visual categories and elements based on LLM-generated outputs may also reflect inherent biases. Although hallucinations and other major errors were rare in our tests, the 17% missing objects underscore LLMs' limitations in accurately capturing complex visual concepts. Future research should systematically benchmark and evaluate these biases in the visual domain and develop improved methodological approaches to pursue annotation accuracy and interpretive reliability. Finally, this study did not assess the audience's perception or interpretation of surreal AIGIs. Future experimental research is needed to examine how audiences perceive and engage with surreal AIGIs and under what conditions surrealism enhances or hinders comprehension, trust, and emotional response.

## Acknowledgments

Appendix of this study and replication materials are available at <https://osf.io/mfwbr/overview?6ccb52f4c2864bb7a33eb67d25dfee05OSF>.

This work was partially supported by the National Science Foundation (CNS-2150716, CNS-2150723).

## References

- AlDahoul, N., Rahwan, T., & Zaki, Y. (2025). Ai-generated faces influence gender stereotypes and racial homogenization. *Scientific Reports*, *15*(1), 14449. <https://doi.org/10.1038/s41598-025-99623-3>
- Arias-Bolzmann, L., Chakraborty, G., & Mowen, J. C. (2000). Effects of absurdity in advertising: The moderating role of product category attitude and the mediating role of cognitive responses. *Journal of Advertising*, *29*(1), 35–49. <https://doi.org/10.1080/00913367.2000.10673602>
- Bandaru, A., Bindley, F., Bluth, T., Chavda, N., Chen, B., & Law, E. (2025). Revealing political bias in llms through structured multi-agent debate. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2506.11825>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM)*. <https://doi.org/10.1609/icwsm.v3il.13937>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. <http://www.jmlr.org/papers/v3/blei03a.html>
- Breton, A. (1969). *Manifestoes of surrealism*. University of Michigan Press.
- Breuer, A., Dietrich, B. J., Crespino, M. H., Butler, M., Pyrsse, J., & Imai, K. (2025). Using ai to summarize us presidential campaign tv advertisement videos, 1952-2012. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2503.22589>
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv preprint*. <https://doi.org/10.48550/ARXIV.2303.04226>
- Chen, Y. T., & Zou, J. (2023). TWIGMA: A dataset of AI-generated images with metadata from Twitter. *Advances in Neural Information Processing Systems*, *36*. <https://arxiv.org/abs/2306.08310>
- Chen, Z., Sun, W., Wu, H., Zhang, Z., Jia, J., Ji, Z., Sun, F., Jui, S., Min, X., Zhai, G., & Zhang, W. (2024). Exploring the naturalness of AI-generated images. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2312.05476>
- Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2306.14924>
- Cozzolino, D., Poggi, G., Corvi, R., Nießner, M., & Verdoliva, L. (2024). Raising the bar of ai-generated image detection with CLIP. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4356–4366. <https://doi.org/10.1109/CVPRW63382.2024.00439>
- Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10850–10869. <https://doi.org/10.1109/TPAMI.2023.3261988>

- Cui, H., Fang, X., Zhang, Z., Xu, R., Kan, X., Liu, X., Yu, Y., Li, M., Song, Y., & Yang, C. (2023). Open Visual Knowledge Extraction via Relation-Oriented Multimodality Model Prompting. *Advances in neural information processing systems*, *36*, 23499–23519. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11315466/>
- Devlin, K., & Cheetham, J. (2023, March). *Fake Trump arrest photos: How to spot an AI-generated image*. BBC. <https://www.bbc.com/news/world-us-canada-65069316>
- DiResta, R., & Goldstein, J. A. (2024). How spammers and scammers leverage AI-generated images on Facebook for audience growth. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2403.12838>
- Farrell, H., Gopnik, A., Shalizi, C., & Evans, J. (2025). Large ai models are cultural and social technologies. *Science*, *387*(6739), 1153–1156. <https://doi.org/10.1126/science.adt9819>
- Ferwerda, J. A. (2003). Three varieties of realism in computer graphics. *Proceedings of SPIE: Human Vision and Electronic Imaging VIII*, 290. <https://doi.org/10.1117/12.473899>
- Gelbrich, K., Gähke, D., & Westjohn, S. A. (2012). Effectiveness of absurdity in advertising across cultures. *Journal of Promotion Management*, *18*(4), 393–413. <https://doi.org/10.1080/10496491.2012.693058>
- Griffin, J. (2025). Re-assessing photorealism: Do we know what we want from art? *The Yale Review*, *113*(1), 161–178. <https://muse.jhu.edu/pub/1/article/953491>
- Growcoot, M. (2023, February). *AI image fools judges and wins photography contest*. PetaPixel. <https://petapixel.com/2023/02/10/ai-image-fools-judges-and-wins-photography-contest/>
- Halperin, B. A., & Lukin, S. M. (2024). Artificial dreams: Surreal visual storytelling as inquiry into AI “hallucination”. *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, 619–637. <https://doi.org/10.1145/3643834.3660685>
- Hausken, L. (2024). Photorealism versus photography. AI-generated depiction in the age of visual disinformation. *Journal of Aesthetics & Culture*, *16*(1), 2340787. <https://doi.org/10.1080/20004214.2024.2340787>
- Homer, P. M., & Kahle, L. R. (1986). A social adaptation explanation of the effects of surrealism on advertising. *Journal of Advertising*, *15*(2), 50–60. <https://doi.org/10.1080/00913367.1986.10673005>
- Joo, J., & Steinert-Threlkeld, Z. C. (2022). Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research*, *4*(1). <https://doi.org/10.5117/CCR2022.1.001.JOO>
- Joon, J. S. (2010). Principles of photorealism to develop photorealistic visualisation for interface design: A review. *Seventh International Conference on Computer Graphics, Imaging and Visualization*, 17–25. <https://doi.org/10.1109/CGIV.2010.12>
- Kamali, N., Nakamura, K., Chatzimparmpas, A., Hullman, J., & Groh, M. (2024). How to distinguish AI-generated images from authentic photographs. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2406.08651>
- Kamali, N., Nakamura, K., Kumar, A., Chatzimparmpas, A., Hullman, J., & Groh, M. (2025). Characterizing photorealism and artifacts in diffusion model-generated

- images. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–26. <https://doi.org/10.1145/3706598.3713962>
- Kavi, A. (2025, May). *Trump denies posting a purported AI image of himself as pope*. The New York Times. <https://www.nytimes.com/2025/05/05/us/politics/trump-pope-photo-ai.html>
- Korkut, U., McGarry, A., Erhart, I., Eslen-Ziya, H., & Jenzen, O. (2022). Looking for truth in absurdity: Humour as community-building and dissidence against authoritarianism. *International Political Science Review*, 43(5), 629–647. <https://doi.org/10.1177/0192512120971151>
- Kratky, A. (2022). Poetic automatism. In *ArtsIT, interactivity and game creation* (pp. 359–378). Springer International Publishing. [https://doi.org/10.1007/978-3-030-95531-1\\_25](https://doi.org/10.1007/978-3-030-95531-1_25)
- Lee, S., Peng, T.-Q., Goldberg, M. H., Rosenthal, S. A., Kotcher, J. E., Maibach, E. W., & Leiserowitz, A. (2024). Can large language models estimate public opinion about global warming? an empirical assessment of algorithmic fidelity and bias. *PLOS Climate*, 3(8), e0000429. <https://doi.org/10.1371/journal.pclm.0000429>
- Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. *Advances in Neural Information Processing Systems*, 36, 34892–34916. [https://proceedings.neurips.cc/paper\\_files/paper/2023/hash/6dcf277ea32ce3288914faf369fe6de0-Abstract-Conference.html](https://proceedings.neurips.cc/paper_files/paper/2023/hash/6dcf277ea32ce3288914faf369fe6de0-Abstract-Conference.html)
- Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X., & Ouyang, W. (2023). Seeing is not always believing: Benchmarking human and model perception of AI-generated images. *Advances in Neural Information Processing Systems*, 36, 25435–25447. <https://arxiv.org/abs/2304.13023>
- Luccioni, S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). Stable bias: Evaluating societal representations in diffusion models. *Advances in Neural Information Processing Systems*, 36, 56338–56351. <https://arxiv.org/abs/2303.11408>
- Lyu, S., & Farid, H. (2005). How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 53(2), 845–850. <https://doi.org/10.1109/TSP.2004.839896>
- Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. *Nature Communications*, 15(1), 654. <https://doi.org/10.1038/s41467-024-44824-z>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., & Häussler, T. (2018). Applying lda topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Menczer, F., Crandall, D., Ahn, Y.-Y., & Kapadia, A. (2023). Addressing the harms of AI-generated inauthentic content. *Nature Machine Intelligence*, 5(7), 679–680. <https://doi.org/10.1038/s42256-023-00690-w>
- Messararis, P. (1996). *Visual persuasion: The role of images in advertising*. SAGE Publications.

- Patel, N. (2024, September). *Let's compare Apple, Google, and Samsung's definitions of a photo*. The Verge. <https://www.theverge.com/2024/9/23/24252231/lets-compare-apple-google-and-samsungs-definitions-of-a-photo>
- Peng, Q., Lu, Y., Peng, Y., Qian, S., Liu, X., & Shen, C. (2025). Crafting synthetic realities: Examining visual realism and misinformation potential of photorealistic AI-generated images. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3706599.3719834>
- Peng, Y., & Lu, Y. (2023). Chapter 4: Computational visual analysis in political communication. In *Research Handbook on Visual Politics*. Edward Elgar Publishing.
- Peng, Y., Qian, S., Lu, Y., & Shen, C. (2025). Large language model-informed feature discovery improves prediction and interpretation of credibility perceptions of visual content. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2504.10878>
- Pereira, L., Roberti Junior, W. C., & Silva, R. L. S. (2021). Photorealism in mixed reality: A systematic literature review. *International Journal of Virtual Reality*, 21(1), 15–29. <https://doi.org/10.20870/IJVR.2021.21.1.3166>
- Pham, C. M., Hoyle, A., Sun, S., Resnik, P., & Iyyer, M. (2024). TopicGPT: A prompt-based topic modeling framework. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2311.01449>
- Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjeh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences*, 121(34), e2308950121. <https://doi.org/10.1073/pnas.2308950121>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Schröter, J. (2023). The ai image, the dream, and the statistical unconscious. *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft*, 19(1), 112–120. <https://doi.org/http://dx.doi.org/10.25969/mediarep/22315>
- Schulze Buschoff, L. M., Akata, E., Bethge, M., & Schulz, E. (2025). Visual cognition in multimodal large language models. *Nature Machine Intelligence*, 7(1), 96–106. <https://doi.org/10.1038/s42256-024-00963-y>
- Serrano, M. Á., Boguñá, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16), 6483–6488. <https://doi.org/10.1073/pnas.0808904106>
- Shahgir, H. S., Sayeed, K. S., Bhattacharjee, A., Ahmad, W. U., Dong, Y., & Shahriyar, R. (2024). IllusionVQA: A challenging optical illusion dataset for vision language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2403.15952>
- Smith, A., & Hutson, J. (2024). Satirical deepfakes, surreal dreamscapes & nostalgic pixels: The rapid evolution and cultural commentary of ai-aesthetics. *ISAR Journal of Arts, Humanities and Social Sciences*, 2(9), 31–43. <https://digitalcommons.lindenwood.edu/faculty-research-papers/676/>

- Smith, N., & Southerton, C. (2025). AI and aesthetic alienation: The image and creativity in contemporary culture. *Social Science Computer Review*, 44(1). <https://doi.org/10.1177/08944393251361449>
- Somepalli, G., Singla, V., Goldblum, M., Geiping, J., & Goldstein, T. (2023). Diffusion art or digital forgery? investigating data replication in diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6048–6058. <https://doi.org/10.48550/arXiv.2212.03860>
- Stuhler, O., Ton, C. D., & Ollion, E. (2025). From codebooks to promptbooks: Extracting information from text with generative large language models. *Sociological Methods & Research*, 54(3), 794–848. <https://doi.org/10.1177/00491241251336794>
- Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2023). Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI. *Journal of Computer-Mediated Communication*, 29(1), zmad045. <https://doi.org/10.1093/jcmc/zmad045>
- Sun, Y., Pendyala, V., Lian, R., Xin, H., Patel, P., Bucy, E. P., & Shah, D. V. (2025). From internet meme to the mainstream: Using computer vision to track “Pepe the Frog” across news platforms. *Visual Communication Quarterly*, 32(1), 33–55. <https://doi.org/10.1080/15551393.2025.2455495>
- The Museum of Modern Art. (1994, March). *American surrealist photography*. The Museum of Modern Art. <https://www.moma.org/calendar/exhibitions/422>
- Walter, D., & Ophir, Y. (2019). News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13(4), 248–266. <https://doi.org/10.1080/19312458.2019.1639145>
- Wasielowski, A. (2024). Unnatural images: On AI-generated photographs. *Critical Inquiry*, 51(1), 1–29. <https://doi.org/10.1086/731729>
- Wei, Y., & Tyson, G. (2024). Understanding the impact of AI-generated content on social media: The Pixiv case. *Proceedings of the 32nd ACM International Conference on Multimedia*, 6813–6822. <https://doi.org/10.1145/3664647.3680631>
- Williams, A. (2023). *Samsung’s moon shots force us to ask: How much AI is too much?* Wired. <https://www.wired.com/story/samsungs-moon-shots-force-us-to-ask-how-much-ai-is-too-much/>
- Wu, J., Gan, W., Chen, Z., Wan, S., & Yu, P. S. (2023). Multimodal large language models: A survey. *2023 IEEE International Conference on Big Data (BigData)*, 2247–2256. <https://doi.org/10.1109/BigData59044.2023.10386743>
- Xiao, K., Yang, L., Tulajiang, P., & Lin, H. (2025). VisualQuest: A diverse image dataset for evaluating visual recognition in LLMs. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2503.19936>