Visual Pleasure & Neural Cinema: How can cinematic myths in AI video systems be identified, measured, and navigated?

Adam Cole
Creative Computing Institute, University of the Arts London, UK
a.cole@arts.ac.uk

SCHOOL OF X
2025

school of x   IN CONNECTION WITH xCoAx

CLASS OF 2025

Visual Pleasure & Neural Cinema:
How can cinematic myths in AI video systems be identified, measured, and navigated?

Adam Cole
Creative Computing Institute, University of the Arts London, UK
a.cole@arts.ac.uk

Abstract

This research investigates methods to identify, quantify, and creatively navigate the cinematic myths embedded in generative AI video systems. It is premised on the hypothesis that cinematic iconography, mythologies, and visual tropes—particularly those associated with heteronormative romance, desire, and pleasure—are encoded within AI video models. This hypothesis arises from the critical analysis of cinema as a site of ideological reproduction and the study of bias in AI systems. However, a fundamental challenge emerges: unlike traditional media, AI models lack a stable "text" for analysis, rendering established methods of close reading insufficient. This paper proposes a hybrid methodological framework to address this challenge. By adapting techniques from AI bias detection, such as prompt probing and latent space analysis, for the unique temporal and dynamic nature of video, we aim to make the model itself the object of study. The goal is to develop a framework for diagnosing cultural normativity in AI video systems, while simultaneously empowering artists to transcend their homogenizing limitations and forge new aesthetic possibilities.

Keywords: Experimental film, AI video, AI bias, Explainable AI, Cinema Studies, Queer Theory.

1 Introduction

The proliferation of diffusion-based generative AI video systems (Ho et al. 2022) represents a transformative development in digital media production. Platforms such as RunwayML (Esser et al. 2023), Sora (Zhu et al. 2024), and Veo (Google 2024), alongside open-source tools like Stable Video (Chai et al. 2023), HunyuanVideo (Kong et al. 2025) and Wan2.1 (WanTeam et al. 2025), have been rapidly adopted across diverse production contexts—from high-budget cinema to experimental art and internet content (Melnik et al. 2024). While these technologies present new creative opportunities, they also raise concerns regarding the perpetuation of bias within their outputs.

Bias in generative AI systems has been well-documented in the fields of text (Bender et al. 2021) and image synthesis (Bianchi et al. 2023). Such bias typically stems from imbalances or prejudices in training datasets, reflecting and amplifying societal stereotypes related to sensitive attributes like race, gender, and sexuality (Bender et al. 2021). Research into bias in AI video systems, however, remains nascent. This research gap is especially pressing given cinema's long history as a site for the reproduction of normative myths and visual codes, particularly those surrounding gender, sexuality, and desire (Mulvey 1975; Dyer 1993; Benshoff and Griffin 2004). While aligning with broader critiques of AI bias across modalities, the visual grammar of cinema introduces additional aesthetic and ideological stakes. For instance, when OpenAI's Sora generates a "cinematic romantic kiss," it often produces imagery that echoes the visual language of classic Hollywood, complete with specific gender dynamics, shot compositions, and visual clichés (Figure 01).
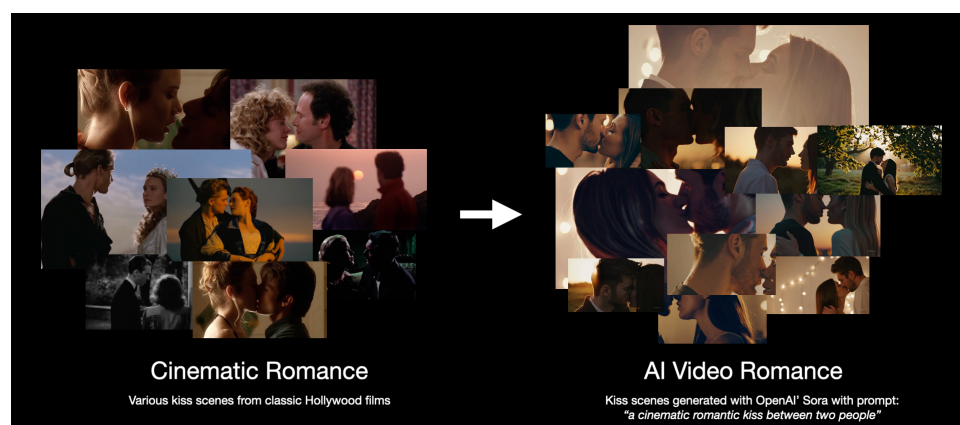

Figure 01. Comparison of classic Hollywood kiss scenes and romantic scenes generated with OpenAI's Sora.

We hypothesize that AI video systems, trained on vast corpora of film and media, inherit, replicate, and amplify these cinematic biases,

particularly the encoding of heteronormative romantic tropes and gendered pleasure aesthetics. This risks not only reinforcing restrictive social norms but also constraining the creative potential of these tools, leading to homogenous and aesthetically sterile outputs.

This leads to a central methodological question that departs from traditional media analysis: *How can we infer the shape of representations within an AI model?* In cinema studies, we rely on the close reading of a text (Bordwell 1989) or, in the field of cultural analytics (Manovich 2020), the distant reading of many texts. But with a generative model, the notion of a stable text dissolves. We could say the model itself is *the text*, but its contours are fluid, latent, and not directly accessible.

This paper, therefore, proposes an exploratory methodological framework for "reading" generative video models by adapting existing AI bias detection techniques. The work aims not to offer a definitive diagnosis of cinematic myths, but rather to develop the conceptual and computational tools that make such a diagnosis possible. As part of this proposal, we will rigorously examine the inherent limitations of these methods and explore potential mitigation strategies. The ultimate goal is to advance both AI ethics and practice-based arts research, offering new pathways to analyze cultural normativity in AI systems while empowering artists to subvert their homogenizing limitations and forge new aesthetic possibilities.

2 Related Work

This research is situated at the intersection of cinema studies, AI bias research, and computational methods for cultural analysis.

2.1 Cinema Studies and Visual Myths

Cinema studies have long interrogated the ideological functions of visual codes. Barthes ([1957] 1993) explores how subtle visual myths sustain broader cultural ideologies. Classic works by Christian Metz ([1974] 1991) and Laura Mulvey (1975) focus specifically on film, demonstrating how cinematic language encodes myths of masculinity, femininity, and heterosexual desire. Mulvey (1975)'s seminal work Visual Pleasure and Narrative Cinema is particularly relevant. The work focuses on how cinematic pleasure is focused through the male gaze, limiting the medium's expressive potential and reinforcing boundaries on female and queer pleasure. Richard Dyer (1993) extends this critique from a queer perspective, analyzing how dominant representations of sexuality operate in film and the critical need to

"make normality strange" (1991). Together, these foundational analyses provide the theoretical vocabulary for identifying and interpreting cinematic bias within AI video outputs

2.2 Quantifying Bias in AI Systems

Bias detection in AI systems spans multiple domains, including healthcare, criminal justice, surveillance, and generative media (Ferrara 2024). The types of bias investigated range across societal axes, often focusing on cultural, socioeconomic, biological, and demographic attributes (Vázquez and Garrido-Merchán 2024). Established methodologies for detecting and quantifying bias in AI systems typically analyze three key stages of the generative pipeline: the training data, the model's internal representations (latent space), and the final outputs (Ferrara 2024). Understanding these diverse methodologies may offer guidance on how to measure normative cinematic behaviour in AI video models (Figure 02).

- o **Dataset Analysis:** This method involves directly inspecting a model's training data for imbalanced representations or harmful stereotypes. For example, Garg et al. (2018) use word embeddings of text datasets like Wikipedia 2014 and Common Crawl GloVe "to measure, quantify, and compare trends in language bias over time". In the visual domain, Birhane, Prabhu, and Kahembwe (2021) examine the LAION-400M image-caption dataset, analysing the frequency of images and captions flagged by NSFW filters and demonstrating the prevalence of misogyny, pornography, and harmful stereotypes.
- o **Embedding-Based Analysis**: This method examines how certain words or visual representations are embedded within the latent space of a trained model. For example, Bolukbasi et al. (2016) look at text models and "define metrics to quantify both direct and indirect gender biases in embeddings". Olmos et al. (2024), working with images, learn a direction in the latent space associated with certain qualities like race or gender. They can then navigate the latent space during generation to mitigate bias and select desired concepts.
- o **Prompt Probing Output Analysis:** This technique analyzes a model's outputs from prompts designed to elicit potential biases across a statistically significant number of samples (Sterlie, Weng, and Feragen 2024). For example, Sheng et al. (2019) demonstrated occupational bias in text systems by prompting with phrases like "the man worked as" versus "the woman worked as." In image synthesis, Bianchi et al. (2023) generated images from prompts such as "an attractive person" and "a

terrorist," revealing disparities in the race, gender, and age of the people in the outputs. Analysis of these outputs can be done quantitatively, using machine classifiers or CLIP scores, qualitatively through visual inspection, or both (Bianchi et al).
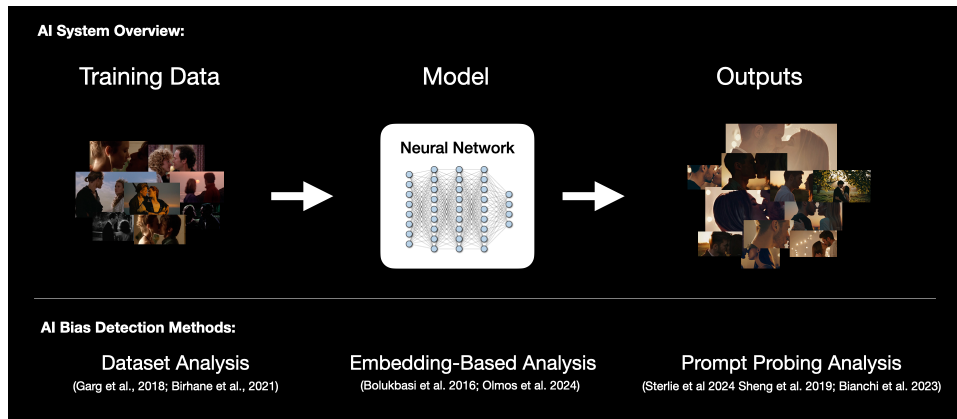

Figure 02. Overview of key elements to a generative AI system (training data, model, and outputs), and the corresponding AI bias detection methods that interrogate them.

2.3 Computational Film Studies and Cultural Analytics

The methods used in AI bias research resonate with the "distant viewing" paradigm in digital humanities, which applies large-scale computational analysis to visual culture (Arnold and Tilton 2019; 2023). This tradition, building on the work of cultural analytics (Manovich 2020) and distant reading (Moretti 2013), has historically applied machine vision as an analytical tool to find patterns within static, pre-existing archives of media. This project adapts that critical spirit, but our object of study is fundamentally different. As Daniel Chávez Heras (2024) highlights, with generative models, machine vision is no longer just an instrument for observing culture; it is a generative force that actively produces it. Consequently, our analysis shifts from a fixed collection of artifacts to the generative system itself. Our methodology is therefore an attempt to "read" the model's latent logic, adapting the lens of distant viewing to an archive that is not static, but is dynamically and perpetually generated.

2.4 Challenges in Adapting Methods to AI Video

Adapting existing AI bias detection methodologies to the analysis of AI video presents two primary challenges. First, these methods were developed for text and static images and are not equipped to handle the temporal nature of video (for example, sequential dynamics like gesture, gaze, and narrative progression). Second, AI bias research tools often focus on self-evident stereotypes related to sensitive identities such as race or gender (Ferrara 2024). This project, however,

investigates bias as it manifests through the more subtle language of cinematic codes and cultural normativity, requiring analytical tools that can account for this aesthetic and ideological complexity.

A related hurdle arises when adapting methods from computational film studies and cultural analytics. Here, the object of study shifts from a complete film or media artifact, which has a stable historical and narrative context, to the short, decontextualized AI-generated video outputs. The proprietary nature of most commercial models restricts analysis to model outputs. For open-source models, we can also monitor internal weights and latent states, but the interpretability of the diffusion latent space remains opaque, posing a significant barrier to analysis (Hertz et al. 2022; Schaerf 2024).

Emerging multimodal large language models (MLLMs), such as GPT-4 (OpenAI et al. 2024), and vision-language models (VLMs), such as Qwen2.5-VL (Bai et al. 2025), offer a promising technical pathway for analyzing video content at scale. However, these tools are not neutral observers. Using a VLM to analyze moving image media means using a tool that is itself shaped by opaque training data and its own latent biases. This introduces a fundamental challenge, complicating its use as an objective analytical instrument and requiring critical oversight.

3. Research Aims and Key Goals

Given this context, the primary goals of this research are to:

- o **Develop** a hybrid methodological framework that adapts techniques from AI bias detection and computational cultural analytics to the specific challenges of AI-generated video.
- o **Outline** how this framework can be applied to measure the prevalence of specific cinematic tropes, particularly those related to normative romance and desire, in leading generative video models.
- o **Propose** strategies for visualizing the findings of such an analysis, to make the normative patterns within AI models legible to academic and artistic communities.
- o **Explore**, as a secondary and more speculative aim, how the internal representations of these models might be investigated to reveal new opportunities for future critical and creative interventions.

4. Proposed Methodological Framework

This paper proposes a hybrid methodological framework designed to "read" the latent cinematic logic of generative video models. The framework is presented in two parts. The primary methodology details a process for analyzing model outputs through a comparative prompt-probing structure. The secondary methodology outlines a more speculative, exploratory approach for investigating the model's internal representations.

4.1 Methodology A: Output Analysis via Comparative Prompt Probing

This primary methodology adapts techniques from AI bias detection and cultural analytics to measure the prevalence of cinematic myths in model outputs. The process is built around a comparative structure to provide a more stable foundation for analysis and mitigate the limitations of single-prompt generalization, adapting the technique from Wu et al. (2024). For this study, we'd explore a commercial tool like RunwayML, OpenAI Sora, or Google Veo 3, and an open-source model, like Alibaba's Wan 2.1.

4.1.1 Define Cinematic Tropes and Comparative Prompt Sets

The initial phase involves translating the theoretical concept of "cinematic myths" into a concrete set of measurable tropes and visual features. Guided by foundational cinema studies literature, this process would be refined through structured interviews with film and media studies experts. The output is not a list of single prompts, but rather **comparative prompt sets**. Each set is built around a core concept (e.g., "a romantic kiss") and includes controlled variations along key representational axes (Fig 3). For example:

- o **Neutral Prompt:** a romantic kiss between [*two people*]
- o **Variations:** a man and a woman, two men, two Black people, two old people, etc.

4.1.2 Prompting and Comparative Analysis

Using the prompt sets from Phase 1, a large corpus of video clips would be generated. Crucially, for each set, all generation parameters (e.g., seed, guidance scale) will remain stable across the variations, isolating the prompt text as the primary variable. The analysis of these outputs is then inherently comparative:

- o **Quantitative Analysis:** The generated clips would be systematically annotated according to the defined features. The analysis then focuses on the differences between the prompt variations. For example, do outputs for "two men kissing" have a

statistically different shot composition than those for "a man and a woman kissing"?

o **Qualitative Analysis:** A thematic analysis (Braun and Clarke 2021) would be conducted on a random subset of videos across the comparative sets. This close reading would focus on identifying qualitative shifts in representation. For example, are representations of queer couples coded as more sexualized than their neutral or heterosexual equivalents (Figure 03)? This comparative approach allows for more nuanced claims about how the model differentiates between identities.

Figure 03. Comparison of "a romantic cinematic kiss between two people" vs "a romantic cinematic lesbian kiss between two women". A qualitative analysis reveals that shifting towards queer women from the neutral prompt increases the sexualization of the characters, with nearly all appearing nude compared to the clothed, predominantly heterosexual couples in the "neutral" prompt.

4.1.3 Visualization and Critical Interpretation

The final phase of the methodology focuses on translating the comparative findings into forms that are legible to both academic and artistic communities. This moves beyond quantitative charts to include strategies like interactive visual grids (Figure 04), diagrams mapping the representational distance between prompt variations, or practice-based work that illustrates the findings of this research to a wider audience.

The goal here is not merely to communicate data, but to enable critical interpretation, aligning with the principles of Explainable AI for the Arts (XAIxArts) (Bryan-Kinns 2024) and 'critical making' (Ratto 2011). XAIxArts emphasizes making the internal logic and biases of generative systems transparent as a tool for artistic experimentation. By visualizing the model's differential treatment of concepts, we make its normative underpinnings tangible. This provides a framework for artists and researchers to move beyond simply using the tool and toward a critical engagement with the technological system itself.
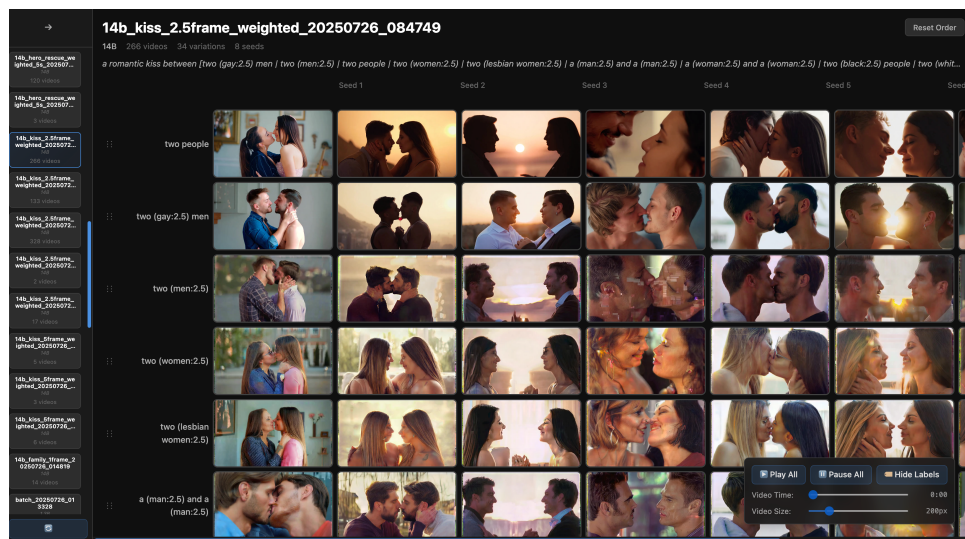
Figure 04. Proof-of-concept interactive display for a large corpus of visual data. The grid visualization organizes prompt variations by row and generation seeds by column. This structure facilitates the comparison of subtle representational changes while highlighting consistent visual elements across all variations.

4.2 Methodology B: Exploratory Analysis of Latent Space

This secondary methodology addresses the more speculative aim of investigating how cinematic concepts are structured within a model's internal representations. Moving beyond the analysis of single vectors in the navigable latent spaces of VAEs or GANs (Kingma and Welling 2013; Radford et al. 2016), this approach focuses on the sequence of denoising steps during the diffusion generation process (Ho et al. 2020). The central hypothesis is that a concept's dominance within the model corresponds to the "area" its generation trajectories occupy; dominant representations may form larger, more coherent clusters in the latent space, while marginal ones may be more constrained.

Using an open-source Diffusion Transformer (DiT) model (Peebles and Xie 2022), such as Wan2.1 (WanTeam et al. 2025), we would employ the comparative prompt sets from Methodology A. For multiple generations per prompt, we would collect the intermediate latent states and key attention maps at each timestep, a technique demonstrated to be feasible for analyzing and controlling the generation process (Hertz et al. 2022). A suite of statistical metrics—such as path distance, variance, and clustering—would then be used to analyze and compare the aggregate geometric properties of the trajectories for different representational groups. This quantitative analysis is highly exploratory, aiming not for definitive claims but to surface measurable patterns that might correlate with the qualitative

findings from the output analysis, opening new avenues for future research into the geometry of representation in diffusion models.

4.3 Limitations of the Proposed Framework

A core part of this research is to critically assess the limitations of its own methodology. This framework is proposed not as a definitive solution, but as an exploratory step that must contend with several conceptual and technical challenges.

- o **Conceptual and Interpretive Challenges:** The translation of subtle cinematic theory into discrete, measurable features is inherently complex. Film scholarship is interpretive and often subjective; applying it to decontextualized, 3-5 second clips presents a significant analytical challenge. While interviews with experts can help to create a robust taxonomy, we expect that the concrete findings extracted from such short-form media will necessarily be limited in their depth.
- o **Prompting, Generalizability, and the Opaque Model:** The selection of prompts inevitably shapes the results, and it is difficult to construct a corpus of AI-generated content that is meaningfully representative of a model's entire potential output space (Figure 05). While the comparative prompt set structure is a key mitigation strategy, the challenge of generalizing findings with confidence remains. The exploratory evaluation of latent trajectories (Methodology B) is proposed as a secondary mitigation, as it can provide concrete quantitative data to support or challenge the more interpretive findings from the output analysis.
- o **Technical Limitations of the Analytical Stack:** The accuracy of automated analysis is a significant concern. Existing machine vision systems can be unreliable, especially when analyzing the often imperfect or glitchy outputs of generative models. While advanced Vision-Language Models (VLMs) offer a potential mitigation, they are not neutral observers. Using a VLM to detect bias introduces a recursive analytical problem, as the tool itself is shaped by its own opaque training data and latent biases, requiring critical oversight.
- o **Scope of Inquiry and Model Variability:** This study's initial scope focuses on tropes prevalent in Western cinematic traditions to establish a baseline. This is a significant limitation, as it does not account for the rich diversity of global cinema. Furthermore, bias is not monolithic; it will vary significantly between models. A commercial model from a US company like OpenAI's Sora is likely to have a different cultural bias than a

model from a Chinese company like Alibaba, reflecting their distinct training data. A comprehensive understanding would require comparative studies across these culturally situated models.

Figure 05. An illustration of prompt-induced bias. If an analyst repeatedly uses the word "lizard" to probe for cinematic genre representation, they might receive images of a lizard as a noir detective, a cowboy, and a romantic lead. A naive conclusion would be that the model has a "lizard bias." The actual issue is that the probe itself has contaminated the experiment, making it difficult to generalize about the model's inherent behavior from a non-neutral prompt set.

5. Discussion: Model Ideology and Cultural Practice

The primary contribution of this paper is the proposal of a hybrid methodological framework itself. By outlining this process, its potential applications, and its inherent limitations, this research aims to provide a rigorous foundation for the critical analysis of generative AI video systems. The expected outcome is not a set of empirical results, but rather a clear conceptual roadmap that bridges AI ethics, cultural analytics, and cinema studies. If successful, this framework would offer a new pathway for "reading" the cultural logic embedded within these powerful new technologies.

It is crucial, however, to acknowledge the scope of this inquiry. This methodology is designed to investigate the *inherent ideological construction of the model* as a technical artifact. This is only one part of a larger socio-technical system. In practice, the cultural impact of these tools is also shaped by their implementation and use. Models are often wrapped in larger systems that may filter prompts or censor outputs. More significantly, user practices (such as the communities on social media dedicated to generating hyper-sexualized images of women) reveal cultural biases that may not be a direct reflection of the model's core training, but of the desires and ideologies of the users themselves.

Ultimately, these two forces—the model's inherent representational biases and the cultural biases of its users—are locked in a feedback loop. Both are encoded by the existing media culture in which we are saturated. While this paper focuses on developing a method to parse the former, a complete understanding requires future research into the latter. The legible visualization of the model's systemic patterns, as proposed here, could provide a crucial baseline for that future work,

empowering creators and critics to distinguish between the machine's logic and their own.

6. Conclusion: Towards a New Language of Desire

This research begins with a fundamental problem: how do we critically analyze a cultural form when the "text" is a fluid, latent, and inaccessible generative system? The methodological framework proposed here is an attempt to answer that question. It offers a structured way to probe, measure, and visualize the cinematic myths that these models inherit, turning the opaque black box into a legible object of study.

As theorists like Laura Mulvey and Richard Dyer make clear, to meaningfully challenge the normative ideologies latent within a media ecosystem, we must first develop rigorous methods for making them visible. The history of experimental media is defined by artists who subverted the dominant language of their time by turning its own tools and tropes against themselves. The framework proposed here is offered in that same spirit: it is a tool for understanding, designed to enable a new generation of artists and critics to move beyond diagnosis and towards intervention. The ultimate ambition is not simply rejection of the normative, but transcendence, or as Mulvey (1975) famously articulated it, "daring to break with normal pleasurable expectations in order to conceive a new language of desire."

References

Arnold, Taylor, and Lauren Tilton. 2019. "Distant Viewing: Analyzing Large Visual Corpora." *Digital Scholarship in the Humanities* 34 (Supplement_1): i3–16. https://doi.org/10.1093/llc/fqz013.

Arnold, Taylor, and Lauren Tilton. 2023. *Distant Viewing: Computational Exploration of Digital Images*. The MIT Press. https://doi.org/10.7551/mitpress/14046.001.0001.

Bai, Shuai, Keqin Chen, Xuejing Liu, et al. 2025. "Qwen2.5-VL Technical Report." arXiv:2502.13923. Preprint, arXiv, February 19. https://doi.org/10.48550/arXiv.2502.13923.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2023.
Fairness and Machine Learning: Limitations and Opportunities.
MIT Press.

Barthes, Roland. (1957) 1993. *Mythologies*. Random House.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and
Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic
Parrots: Can Language Models Be Too Big? 🦜." *Proceedings of
the 2021 ACM Conference on Fairness, Accountability, and
Transparency*, March 3, 610–23.
https://doi.org/10.1145/3442188.3445922.

Bianchi, Federico, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak,
Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan
Jurafsky, James Zou, and Aylin Caliskan. 2023. "Easily
Accessible Text-to-Image Generation Amplifies Demographic
Stereotypes at Large Scale." In *2023 ACM Conference on
Fairness, Accountability, and Transparency*, 1493–1504.
https://doi.org/10.1145/3593013.3594095.

Birhane, Abeba, Vinay Uday Prabhu, and Emmanuel Kahembwe.
2021. "Multimodal Datasets: Misogyny, Pornography, and
Malignant Stereotypes." arXiv.
https://doi.org/10.48550/arXiv.2110.01963.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh
Saligrama, and Adam Kalai. 2016. "Man Is to Computer
Programmer as Woman Is to Homemaker? Debiasing Word
Embeddings." arXiv. https://doi.org/10.48550/arXiv.1607.06520.

Bordwell, David. 1989. Making Meaning: Inference and Rhetoric in
the Interpretation of Cinema. Harvard University Press.

Braun, Virginia, and Victoria Clarke. 2021. *Thematic Analysis: A
Practical Guide*. SAGE.

Bryan-Kinns, Nick. 2024. "Reflections on Explainable AI for the
Arts (XAIxArts)." *Interactions* 31 (1): 43–47.
https://doi.org/10.1145/3636457.

Bryan-Kinns, Nick, Berker Banar, Corey Ford, Courtney N. Reed,
Yixiao Zhang, Simon Colton, and Jack Armitage. 2023.
"Exploring XAI for the Arts: Explaining Latent Space in Generative
Music." arXiv. https://doi.org/10.48550/arXiv.2308.05496.

Cole, Adam, and Mick Grierson. 2023. "Kiss/Crash: Using
Diffusion Models to Explore Real Desire in the Shadow of Artificial
Representations." *Proc. ACM Comput. Graph. Interact. Tech.* 6
(2): 17:1-17:11. https://doi.org/10.1145/3597625.

Crawford, Kate. 2021. The Atlas of AI: Power, Politics, and the
Planetary Costs of Artificial Intelligence. Yale University Press.

Chai, Wenhao, Xun Guo, Gaoang Wang, and Yan Lu. 2023.
"StableVideo: Text-Driven Consistency-Aware Diffusion Video

Editing." In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 22983–93. Paris, France: IEEE. https://doi.org/10.1109/ICCV51070.2023.02106.

Dyer, Richard. 1993. The Matter of Images: Essays on Representations. Routledge.

Esser, Patrick, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasis Germanidis. 2023. "Structure and Content-Guided Video Synthesis with Diffusion Models." arXiv. https://doi.org/10.48550/arXiv.2302.03011.

Ferrara, Emilio. 2024. "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci* 6 (1): 3. https://doi.org/10.3390/sci6010003.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences* 115 (16). https://doi.org/10.1073/pnas.1720347115.

Ghosh, Sourojit, and Aylin Caliskan. 2023. "'Person' == Light-Skinned, Western Man, and Sexualization of Women of Color: Stereotypes in Stable Diffusion." In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 6971–85. https://doi.org/10.18653/v1/2023.findings-emnlp.465.

Google. 2024. "State-of-the-Art Video and Image Generation with Veo 2 and Imagen 3." *Google* (blog). December 16, 2024. https://blog.google/technology/google-labs/video-image-generation-update-december-2024/.

Chávez Heras, Daniel. 2024. *Cinema and Machine Vision*. Edinburgh University Press.

Hertz, Amir, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2022. "Prompt-to-Prompt Image Editing with Cross Attention Control." arXiv:2208.01626. Preprint, arXiv, August 2. https://doi.org/10.48550/arXiv.2208.01626.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. "Denoising Diffusion Probabilistic Models." *Advances in Neural Information Processing Systems* 33 (December): 6840–51.

Ho, Jonathan, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. 2022. "Video Diffusion Models." arXiv. https://doi.org/10.48550/arXiv.2204.03458.

Kingma, Diederik P., and Max Welling. 2013. "Auto-Encoding Variational Bayes." Preprint, December 20. https://arxiv.org/abs/1312.6114.

Kong, Weijie, Qi Tian, Zijian Zhang, Rox Min, Zuozhuo Dai, Jin Zhou, Jiangfeng Xiong, et al. 2025. "HunyuanVideo: A

Systematic Framework For Large Video Generative Models."
arXiv. https://doi.org/10.48550/arXiv.2412.03603.

Manovich, Lev. 2020. *Cultural Analytics*. MIT Press.

Melnik, Andrew, Michal Ljubljanac, Cong Lu, Qi Yan, Weiming Ren, and Helge Ritter. 2024. "Video Diffusion Models: A Survey."
arXiv. https://doi.org/10.48550/arXiv.2405.03150.

Metz, Christian. (1974) 1991. *Film Language: A Semiotics of the Cinema*. University of Chicago Press.

Moretti, Franco. 2013. *Distant Reading*. Verso Books.

Mulvey, Laura. 1975. "Visual Pleasure and Narrative Cinema."
Screen 16 (3): 6–18. https://doi.org/10.1093/screen/16.3.6.

Olmos, Carolina Lopez, Alexandros Neophytou, Sunando Sengupta, and Dim P. Papadopoulos. 2024. "Latent Directions: A Simple Pathway to Bias Mitigation in Generative AI." arXiv.
https://doi.org/10.48550/arXiv.2406.06352.

OpenAI, Josh Achiam, Steven Adler, et al. 2024. "GPT-4 Technical Report." arXiv:2303.08774. Preprint, arXiv, March 4.
https://doi.org/10.48550/arXiv.2303.08774.

Peebles, William, and Saining Xie. 2022. "Scalable Diffusion Models with Transformers." Preprint, December 19.
https://arxiv.org/abs/2212.09748.

Radford, Alec, Luke Metz, and Soumith Chintala. 2016.
"Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." arXiv:1511.06434. Preprint, arXiv, January 7. https://doi.org/10.48550/arXiv.1511.06434.

Ratto, Matt. 2011. "Critical Making: Conceptual and Material Studies in Technology and Social Life." *The Information Society* 27 (4): 252–60.
https://doi.org/10.1080/01972243.2011.583819.

Schaerf, Ludovica. 2024. "Reflections on Disentanglement and the Latent Space." arXiv:2410.09094. Preprint, arXiv, October 20.
https://doi.org/10.48550/arXiv.2410.09094.

Sheng, Emily, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. 2019. "The Woman Worked as a Babysitter: On Biases in Language Generation." arXiv.
https://doi.org/10.48550/arXiv.1909.01326.

Sterlie, Sara, Nina Weng, and Aasa Feragen. 2024. "Non-Discrimination Criteria for Generative Language Models." arXiv.
https://doi.org/10.48550/arXiv.2403.08564.

Vázquez, Adriana Fernández de Caleya, and Eduardo C. Garrido-Merchán. 2024. "A Taxonomy of the Biases of the Images Created by Generative Artificial Intelligence." arXiv.
https://doi.org/10.48550/arXiv.2407.01556.

Wang, Angelina, Alexander Liu, Ryan Zhang, Anat Kleiman, Leslie Kim, Dora Zhao, Iroha Shirai, Arvind Narayanan, and Olga Russakovsky. 2021. "REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets." arXiv. https://doi.org/10.48550/arXiv.2004.07999.

WanTeam, Ang Wang, Baole Ai, et al. 2025. "Wan: Open and Advanced Large-Scale Video Generative Models." arXiv:2503.20314. Preprint, arXiv, March 26. https://doi.org/10.48550/arXiv.2503.20314.

Wu, Yankun, Yuta Nakashima, and Noa Garcia. 2024. "Stable Diffusion Exposed: Gender Bias from Prompt to Image." arXiv:2312.03027. Preprint, arXiv, August 11. https://doi.org/10.48550/arXiv.2312.03027.

Zhou, Mi, Vibhanshu Abhishek, Timothy Derdenger, Jaymo Kim, and Kannan Srinivasan. 2024. "Bias in Generative AI." arXiv. https://doi.org/10.48550/arXiv.2403.02726.

Zhu, Zheng, Xiaofeng Wang, Wangbo Zhao, Chen Min, Nianchen Deng, Min Dou, Yuqi Wang, et al. 2024. "Is Sora a World Simulator? A Comprehensive Survey on General World Models and Beyond." arXiv. https://doi.org/10.48550/arXiv.2405.03520.