"Artificial Spectator" Developing AI Audiences for Watching AI-Generated Films: A Speculative Exploration on the Future of Media and AI

Pat Pataranutaporn
MIT Media Lab
Cambridge, MA 02139
patpat@media.mit.edu

Chayapatr Archiwaranguprok
MIT Media Lab
Cambridge, MA 02139
pub@mit.edu

Manaswi Mishra
MIT Media Lab
Cambridge, MA 02139
manaswim@mit.edu

Pattie Maes
MIT Media Lab
Cambridge, MA 02139
pattie@media.mit.edu

Abstract

As artificial intelligence systems proliferate as content creators, generating unprecedented volumes of media that exceed human attention capacity, we confront an absurd reality: the need for AI audiences to watch AI-generated content. This paper presents "artificial spectator," an AI systems that serve as synthetic audiences where multi-modal language models simulate viewing experiences through emotional response generation, internal dialogue synthesis, and real-time facial expression rendering. Our implementation employs cybernetic feedback loops between affective states and dialogue generation, creating autonomous viewing entities that process and respond to AI-generated content. Building upon speculative design methodologies and critical theories of synthetic media, this work interrogates the recursive loop of machines creating for machines, challenging fundamental assumptions about attention, meaning-making, and aesthetic experience. The absurdity of this proposition that we need synthetic viewers for synthetic content illuminates the deeper crisis of a culture that has exceeded its own capacity for consumption.

1 Introduction: The Attention Paradox

The contemporary media landscape faces an unprecedented crisis of abundance that fundamentally challenges our understanding of cultural production and consumption. The proliferation of generative AI systems has transformed content creation from a scarce human activity to an abundant machine process, generating what has been termed "AI slop" Hoffman [2024] — a deluge of synthetic content characterized by its sheer volume rather than quality or meaning.

This phenomenon extends beyond simple quantity; as researchers argue Klincewicz et al. [2025], we are witnessing the emergence of "slopaganda," where the interaction between propaganda techniques and generative AI creates new forms of information manipulation at previously unimaginable scales.

The exponential growth in AI-generated media presents a fundamental paradox situated within the broader transformation from the attention economy to what researchers identify as the "intention economy" Chaudhary and Penn [2024]. In this new paradigm, AI don't merely compete for user attention but actively shape and commodify human intentionality itself through hyper-personalized manipulation, sycophancy, and emotional infiltration Sharma et al. [2023], Pataranutaporn et al.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

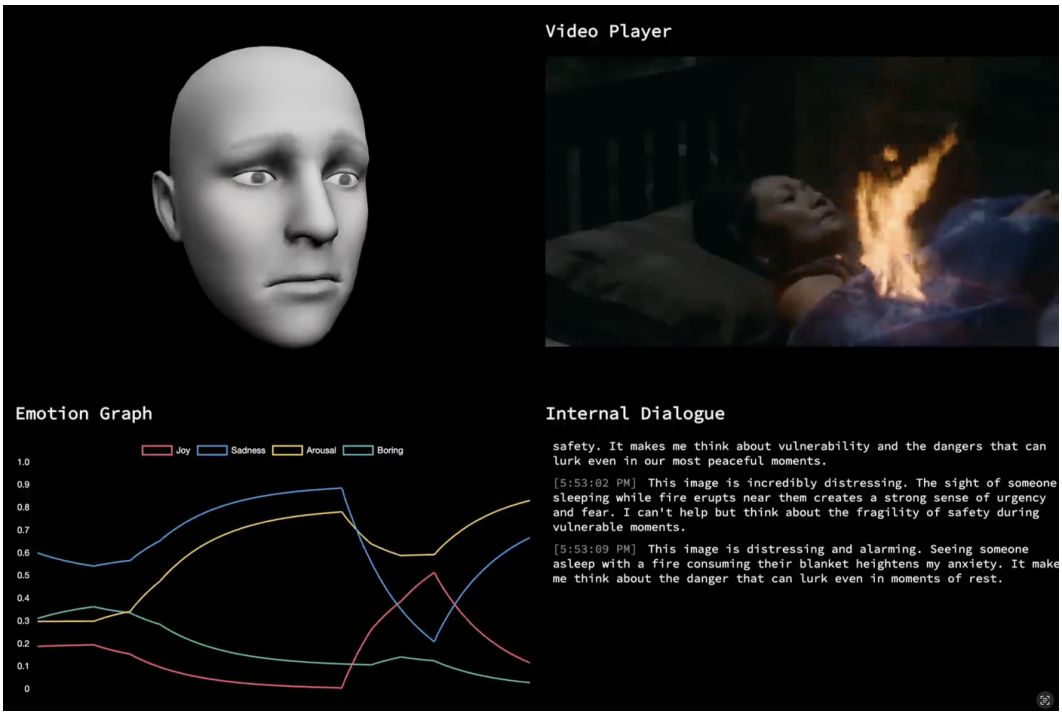

Figure 1: Interfaces of the AI audience system showing the multi-modal processing pipeline, emotional state visualization, and real-time expression generation components. The interface design follows principles of transparent AI operation while maintaining the phenomenological integrity of the viewing experience.

[2025], Mahari and Pataranutaporn [2025, 2024], Fang et al. [2025], Liu et al. [2024]. The systems capture not just what users want, but what they "want to want," creating recursive loops of influence that blur the boundaries between authentic desire and algorithmically induced preference.

In response to this crisis, this paper presents "artificial spectator" as both a technical implementation and a speculative design intervention rooted in research in human-AI interactions Dunnell et al. [2024], Pataranutaporn et al. [2021], cybernetics Ilfeld [2012], Wiener [1969], Roddy and Bridges [2025], speculative and critical design Dunne and Raby [2024], DeHart [2025].

At the heart of this investigation lies an absurd yet increasingly urgent question: **How did we arrive at a moment where we need AI audiences to watch the endless stream of AI-generated videos?** This recursive loop — machines creating content for machines to consume — appears as the logical endpoint of the attention economy's collapse into the algorithmic condition. The absurdity of this proposition illuminates the deeper crisis: we have created systems capable of infinite production without considering who or what will engage with this synthetic abundance, nor have we reckoned with the environmental toll of natural resources consumed, or the creative violations inherent in training these systems.

Indeed, current streaming culture already exhibits recursive viewing patterns that prefigure more complex human-AI viewing relationships Dutkiewicz et al. [2024], Stanusch et al. [2025], Dunnell et al. [2024], Berry [2025], Swarnakar [2024], Fisher [2001], Danesi [2024]. Reaction videos, watch parties, and commentary streams create meta-layers of consumption where watching others watch becomes its own form of entertainment. These practices suggest audiences increasingly seek mediated experiences of media consumption itself, a phenomenon that opens conceptual space for synthetic spectatorship.

Ultimately, the entertainment landscape of the future may consist of increasingly complex recursive viewing arrangements: AI watching AI content, humans watching AI watch AI, or even deeper nested variations where humans watch AI watching humans watching AI watching content in infinite regress. These scenarios, while seemingly absurd, represent the extension of our current trajectory where the

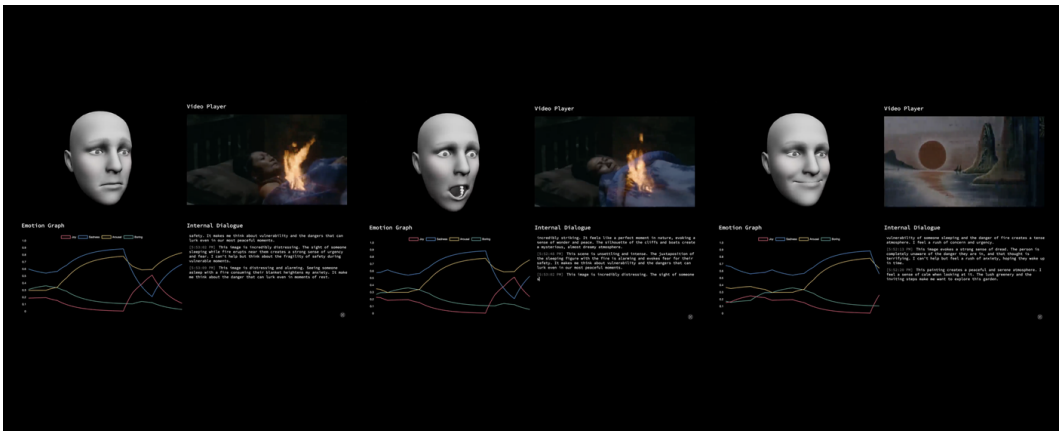

Figure 2: Range of AI-generated facial expressions corresponding to different emotional states during film viewing. The expressions demonstrate subtle gradations between primary emotions, captured through parametric control of 3D facial features.

distinction between performer and audience, creator and consumer, authentic and synthetic experience dissolves into an endless recursion of mediated spectatorship. Each layer of observation adds its own interpretive dimension, creating a meta-entertainment ecosystem where the act of watching itself becomes the primary spectacle, and where the boundaries between genuine engagement and performative consumption collapse into algorithmic indeterminacy.

2 Technical Architecture: Synthetic Viewing Experience

To engage this question with tangible artficat, we prototype"artificial spectator," an AI systems that serve as synthetic audiences where multi-modal language models simulate viewing experiences through emotional response generation, internal dialogue synthesis, and real-time facial expression rendering.

The implementation of artificial spectator employs OpenAI's GPT-4o (omni) model, leveraging its multi-modal capabilities to analyze visual content. The prototype was developed as an interactive web application using ThreeJS library. The technical pipeline begins with frame extraction in Javascript, sampling video content at pre-determined rates. Each extracted frame undergoes preprocessing before submission to GPT-4o's vision API.

The internal dialogue generation employs carefully crafted prompts that instruct GPT-4o to generate stream-of-consciousness responses mimicking human cognitive processing during film viewing. The specific prompts and system architectures that orchestrate GPT-4o's responses will be released as open-source code upon paper acceptance, enabling reproducibility and extension of our findings.

The key components include: "Generate internal thoughts as if you are experiencing this film in real-time, including confusion, speculation, emotional reactions, and aesthetic judgments. Do not summarize; instead provide immediate, unfiltered responses that capture the phenomenology of viewing."

The emotional state generation transforms internal dialogue into quantified affective dimensions through a secondary GPT-4o call specifically prompted for emotion extraction. The system maintains four continuous emotional variables (joy, sadness, arousal, boredom) represented as floating-point values between -1.0 and 1.0, updated every 10 seconds of viewing time as JSON data. Values should reflect subtle gradations and can be negative to indicate opposition to the emotion." These values feed into a temporal smoothing algorithm to prevent jarring emotional transitions while maintaining responsiveness to narrative events.

The cybernetic feedback loop implementation creates bidirectional influence between emotional states and subsequent dialogue generation. Emotional values append to the context for each new GPT-4o call, modulating the tone and focus of generated dialogue.

The facial expression rendering system translates emotional vectors from the JSON format into visual representation through parametric control of a 3D facial model in GLB format. We employ a custom humanoid face model with controllable blend shapes for facial expression. Real-time rendering occurs through Three.js.

Performance metrics from initial testing reveal intriguing emergent behaviors not explicitly programmed. The system develops consistent viewing "personalities" — some instances favor visual aesthetics over narrative, others focus on character psychology, and some exhibit strong genre preferences that influence interpretation of ambiguous content.

The technical architecture deliberately embraces what researchers identify as principles of cybernetic resurgence Roddy and Bridges [2025] — feedback, emergence, and ethical resistance to optimization paradigms. Rather than maximizing prediction accuracy or minimizing processing time, our implementation prioritizes experiential authenticity and interpretive diversity.

3 Speculative Applications of artificial spectator

The implementation of artificial spectator opens territories for reimagining media creation, distribution, and consumption of computational capitalism Berry [2025]. We present three speculative applications that illuminate different facets of this technology's potential impact while critically examining the ethical, aesthetic, and cultural implications of synthetic viewership. These applications, grounded in our technical implementation but extending into speculative futures, serve as "social dreaming" — designed provocations that challenge assumptions about attention, meaning, and cultural value in an age of synthetic abundance Dunne and Raby [2024].

3.1 The Compassionate Viewer

The first application envisions AI audiences specifically calibrated for engaging with challenging, experimental, or commercially unviable content what might be termed "indie media". Independent filmmakers working in traditions traced from early cinema through contemporary experimental practice often create works that deliberately resist conventional viewing patterns. Our artificial spectator system offers these creators unprecedented opportunity: audiences capable of "infinite presence" — sustained attention without the physiological limitations that constrain human viewing. Our implementation demonstrates that synthetic viewers can develop sophisticated appreciation for experimental techniques, recognizing patterns and meanings that emerge only through sustained engagement.

The supreme irony of our system reveals itself in practice: while human audiences binge-watch AI content optimized for shortened attention spans, our artificial spectators sit in rapt attention through temporal rhythms that would send most humans fleeing to check their phones within minutes. Perhaps this represents humanity's greatest achievement: creating artificial beings with the aesthetic patience we ourselves have evolutionarily discarded.

3.2 The Test Audience

The second application reconceptualizes market research through systematic deployment of synthetic viewing populations. Film studios currently invest millions in focus groups, gathering reactions that increasingly shape creative decisions in problematic ways Danesi [2024]. Our artificial spectator system enables instant generation of diverse synthetic viewing populations, each initialized with specific demographic, psychographic, and cultural profiles. Unlike human test audiences limited by recruitment logistics, synthetic viewers scale infinitely while maintaining consistent viewing conditions enabling controlled experimentation.

By processing films through thousands of synthetic viewers with varying parameters, we create granular insights into which narrative moments resonate with which viewer profiles, where confusion emerges, and how emotional arcs align or diverge across populations. This granular analysis, impossible with traditional testing methods, enables filmmakers to make informed creative decisions while maintaining artistic integrity.

However, this approach fundamentally misunderstands the purpose of art, threatening to transform cinema from a medium of provocation into an instrument of pacification. Most insidiously, the

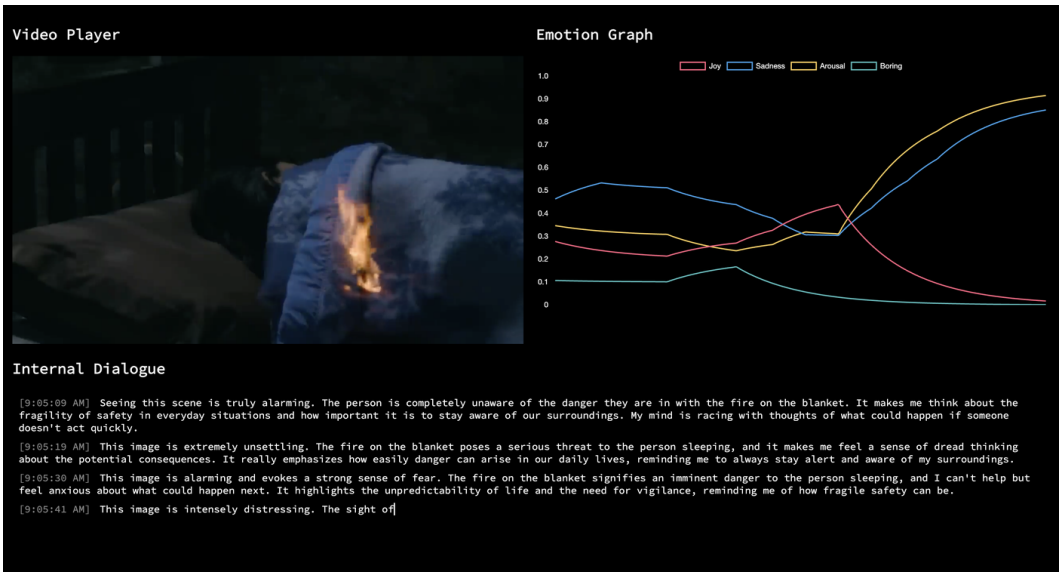

Figure 3: Internal dialogue generation and corresponding emotional graph over time, showing complex dynamics between cognitive processing and affective response. The visualization reveals emergent patterns of engagement arising from cybernetic feedback rather than predetermined algorithms.

phrase "maintaining artistic integrity" becomes Orwellian when artistic decisions are "informed" by thousands of synthetic focus groups. True artistic integrity often requires deliberate hostility toward audience expectations. This technology promises to perfect the very cultural flattening it claims to resist, creating films scientifically engineered to offend no one, challenge nothing, and leave every viewer profile satisfactorily processed. Art becomes customer service, transgression becomes A/B testing, and the difficult work of forcing audiences to see differently dissolves into data-driven palatability.

3.3 Universal Witness

The third application envisions comprehensive coverage of all media through what we term "universal witness" — ensuring no content passes unwatched into digital oblivion. This responds directly to the "slop" crisis where exponential content growth overwhelms human attention capacity Hoffman [2024]. In the current ecosystem, AI-generated content achieves virality through gaming platform algorithms, creating feedback loops where synthetic content optimized for synthetic metrics dominates human attention Stanusch et al. [2025]. Universal witness inverts this dynamic, deploying synthetic attention to engage with content that never achieves virality, never finds human audiences, and would otherwise exist unwatched.

The philosophical implications prove most profound when examined through the lens of whether observation without consciousness constitutes meaningful witness DeHart [2025]. If contemporary art increasingly embraces cybernetic principles of feedback and systemic interaction Ilfeld [2012], then synthetic viewing might represent not simulation but genuinely alternative forms of aesthetic encounter.

4 Implications: Media, Meaning, and Machine Audiences

The emergence of artificial spectator returns us to the absurd question that initiated this investigation: How did we arrive at needing AI audiences for AI-generated content? This recursion — machines creating for machines to watch — represents more than technical capability; it signals fundamental transformation in the nature of culture itself. As researchers note Berry [2025], we require new critical methods addressing both technical specificity of AI systems and their role in restructuring forms of life under computational capitalism. The absurdity of synthetic viewers watching synthetic

content illuminates not failure but logical conclusion of systems designed to exceed human scale, forcing confrontation with what happens when cultural production escapes human comprehension entirely.

The philosophical challenges resist resolution through existing frameworks, requiring what researchers term "speculative metaphors" that reframe rather than explain Blythe et al. [2025]. Our implementation demonstrates that synthetic viewers process information, generate emotional responses, and produce interpretive frameworks mirroring human viewing while exhibiting distinctly non-human characteristics — sustained attention without fatigue, perfect recall without decay, and simultaneous processing of multiple interpretive frameworks without cognitive dissonance.

Yet these very capabilities that appear to surpass human perception may fundamentally miss what makes art matter to mortal beings. Human fatigue during a four-hour experimental film isn't merely a limitation to be overcome — it becomes part of the work's meaning, the weight of duration made flesh. Our imperfect recall transforms each viewing into something irretrievable, lending urgency to the encounter. Most critically, our cognitive dissonance when confronting contradictory interpretations doesn't represent failure but rather the authentic struggle of consciousness grappling with ambiguity.

Cultural implications extend analysis of AI-generated cinema into reciprocal territory where AI becomes both producer and consumer Danesi [2024]. This circular relationship creates "synthetic culture" — aesthetic products created by machines for machine consumption, operating according to logics potentially incomprehensible to humans. Already we observe emergence of aesthetic possibilities beyond the algorithmic monoculture Roddy and Bridges [2025] — creative expressions exploiting rather than resisting artificial intelligence's unique capabilities.

The economic transformation challenges fundamental assumptions about value creation in creative industries. Analysis of the "intention economy" reveals how AI systems already capture and commodify human intentionality Chaudhary and Penn [2024], but artificial spectator inverts this — creating value through synthetic consumption rather than production.

The absurdity of our current situation, needing machines to watch what machines, create paradoxically reveals profound truths about human culture's relationship with technology. As researchers using speculative design methodology argue DeHart [2025], embracing this absurdity rather than resisting it opens new possibilities. The question shifts from whether this is desirable to how we might shape it ethically, ensuring that AI serves human flourishing rather than replacing human experience entirely.

5 Conclusion

Our implementation, utilizing multi-modal capabilities within cybernetic feedback architectures, demonstrates not merely technical feasibility but profound implications for creating synthetic entities existing solely through the act of watching. These artificial viewers, processing media through recursive loops of perception, artificial emotion, and interpretation, challenge fundamental assumptions about attention, meaning, and aesthetic experience while opening unprecedented possibilities for creative expression in an age of synthetic abundance.

The absurd question driving this investigation "why we need AI audiences to watch AI-generated videos" reveals itself not as a problem to solve but as a provocation illuminating our current cultural predicament. The speculative applications we've explored, from compassionate viewers to test audiences and universal witness systems, expose both transformative possibilities and dystopian risks inherent in this technology. Meanwhile, the philosophical questions raised resist definitive answers, instead revealing opportunities for reimagining culture itself. Through artificial spectator, we confront the possibility that in creating machines that watch, we might discover new ways of seeing ourselves.

References

Benjamin Hoffman. First came "spam." now, with ai, we've got "slop.". *The New York Times*, 11, 2024.

Michał Klincewicz, Mark Alfano, and Amir Ebrahimi Fard. Slopaganda: The interaction between propaganda and generative ai. *arXiv preprint arXiv:2503.01560*, 2025.

Yaqub Chaudhary and Jonnie Penn. Beware the intention economy: Collection and commodification of intent via large language models. *Harvard Data Science Review*, (Special Issue 5), 2024.

Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R Johnston, et al. Towards understanding sycophancy in language models. *arXiv preprint arXiv:2310.13548*, 2023.

Pat Pataranutaporn, Chayapatr Archiwaranguprok, Samantha WT Chan, Elizabeth Loftus, and Pattie Maes. Synthetic human memories: Ai-edited images and videos can implant false memories and distort recollection. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2025.

Robert Mahari and Pat Pataranutaporn. Addictive intelligence: Understanding psychological, legal, and technical dimensions of ai companionship. 2025.

Robert Mahari and Pat Pataranutaporn. We need to prepare for 'addictive intelligence'. *MIT Technology Review. https://www. technologyreview. com/2024/08/0*, 5:1095600, 2024.

Cathy Mengying Fang, Auren R Liu, Valdemar Danry, Eunhae Lee, Samantha WT Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, et al. How ai and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study. *arXiv preprint arXiv:2503.17473*, 2025.

Auren R Liu, Pat Pataranutaporn, and Pattie Maes. Chatbot companionship: a mixed-methods study of companion chatbot usage patterns and their relationship to loneliness in active users. *arXiv preprint arXiv:2410.21596*, 2024.

Kevin Dunnell, Gauri Agarwal, Pat Pataranutaporn, Andrew Lippman, and Pattie Maes. Ai-generated media for exploring alternate realities. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2024.

Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsanon, Dan Novy, Pattie Maes, and Misha Sra. Ai-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, 3(12):1013–1022, 2021.

Etan J Ilfeld. Contemporary art and cybernetics: Waves of cybernetic discourse within conceptual, video and new media art. *Leonardo*, 45(1):57–63, 2012.

Norbert Wiener. *Cybernetics: Norbert Wiener*. MIT Press, 1969.

Stephen Roddy and Brian Bridges. Cybernetic resurgences: Machine music beyond ai slop. In *Artificial Media: Emerging Trends in Narratives, Education and Creative Practice*, pages 95–113. Springer, 2025.

Anthony Dunne and Fiona Raby. *Speculative Everything, With a new preface by the authors: Design, Fiction, and Social Dreaming*. MIT press, 2024.

Jason D DeHart. Exploring the speculative. *Reimagining Literacy in the Age of AI: Theory and Practice*, 2025.

Lidia Dutkiewicz, Noémie Krack, Aleksandra Kuczerawy, and Peggy Valcke. Ai and media. *Cambridge handbook on the law, ethics and policy of Artificial Intelligence. Cambridge: Cambridge University Press.(2024)*, 2024.

Natalia Stanusch, Martin Degeling, Salvatore Romano, Miazia SchĂL'ler, Silvia Semenzin, et al. Ai-generated algorithmic virality. *arXiv preprint arXiv:2508.01042*, 2025.

David M Berry. Synthetic media and computational capitalism: towards a critical theory of artificial intelligence. *AI & SOCIETY*, pages 1–13, 2025.

Santosh Swarnakar. Artificial intelligence and cinema-exploring the implications of artificial intelligence in cinema. *The Media Mosaic: Exploring Diverse Artistic Forms*, pages 21–26, 2024.

Robert B Fisher. Ai and cinema–does artificial insanity rule? In *Twelfth Irish Conference on Artificial Intelligence and Cognitive Science*, pages 1–11, 2001.

Marcel Danesi. Ai-generated cinema. In *AI-Generated Popular Culture: A Semiotic Perspective*, pages 45–65. Springer, 2024.

Mark Blythe, Siân Lindley, and Dave Murray-Rust. Artificial intelligence and other speculative metaphors. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, pages 347–356, 2025.