

BEYOND EFFICIENCY: REFRAMING AI HALLUCINATION AND REFUSAL AS AESTHETIC-MORAL GESTURES IN CINEMATIC PRACTICE

IN SUN JUNG

Kunsan National University / Republic of Korea

DOKYUN LEE

Korea University of Media Arts, Republic of Korea

SANGMI IM

Korea University of Media Arts, Republic of Korea

Jung In Sun is a film scholar exploring the intersection of cinema and emerging technologies. Jung's research examines the application of new media in film production, critically investigating how these technological innovations reshape both the creative. ORCID: 0009-0003-2019-0745

Dokyun Lee, Ph.D. in Film and Visual Arts (Kyonggi University), is a Professor in the Department of Film Production at Korea University of Media Arts. His research interests include video art, AI education, and film theory. ORCID: 0009-0000-1448-6783

Sangmi Im, a Ph.D. candidate in Technology Art at the Graduate School of Advanced Imaging Science, Multimedia & Film, Chung-Ang University, is an Assistant Professor in the Department of Game Content at Korea University of Media Arts. Her research interests include new media art and AI art. ORCID: 0009-0008-7045-5656

Corresponding author

In Sun JUNG
Kunsan National University
558, Daehak-ro, Gunsan-si, Jeonbuk-do, Republic of Korea, 54150
cestlavie@kunsan.ac.kr

Acknowledgements

This research was supported by the 2026 Academic Research Grant of Korea University of Media Arts. An earlier version of this paper was presented at the 2025 CILECT Congress. I would like to express my sincere gratitude to everyone who contributed to the development of this research.

Schedule for publication

Paper Submitted: 15th October 2025
Accepted for Publication: 1st April 2026
Published online: 14th April 2026

Abstract

Generative artificial intelligence (AI) in cinematic practice is predominantly analysed through the lens of technical efficiency. This study proposes a critical reframing of algorithmic anomalies in film production, and its aesthetic and moral dimensions. Specifically, the research reframes AI hallucination and refusal, often dismissed as technical flaws, as aesthetic-moral gestures. The study conducts a textual analysis of AI-generated student short films, and applies cognitive and aesthetic-philosophical frameworks from Boden, Levinas, and Sontag. This study proposes a conceptual framework and provides illustrative case studies to explore how emerging creators negotiate cinematic authorship through these algorithmic behaviours. Hallucination is examined as a "creative deviation" that may prompt transformational curation within specific production contexts, where unpredictable visual anomalies are reinterpreted by creators as signs of memory, fragility, or affective ambiguity rather than as direct evidence of audience reception. Conversely, algorithmic refusal is examined as an "ethical pause" and a modern cinematic ellipsis. By restricting explicit depiction, this refusal prompted the selected creators to reconsider subjective moral boundaries and employ symbolic omission and hybrid workflows. Within these pedagogical contexts, hallucination and refusal form a dialectic of deviation and restraint. Generative AI is conceptualised not just as a tool for mechanical replication, but as an interactive catalyst expanding interpretive possibilities. Machine error and algorithmic limitation are thus recontextualised as opportunities for localised human inquiry and moral creativity.

Keywords: *Generative AI, Hallucination, Refusal, Cinematic ellipsis, Transformational creativity, Aesthetic-Moral gestures*

Introduction

Contextualising Generative AI in Cinematic Production

"I can't watch this stuff and find [it] interesting. Whoever creates this stuff has no idea what pain is whatsoever. I am utterly disgusted. If you really want to make creepy stuff, you can go ahead and do it. I would never wish to incorporate this technology into my work at all. I strongly feel that this is an insult to life itself."

— Hayao Miyazaki (Arakawa, 2016, 59:01)

When Miyazaki uttered these words after viewing an AI-generated animation, his visceral reaction became emblematic of a broader cultural unease toward AI in art. His rejection stemmed not only from aesthetic distaste but from the perceived absence of lived experience—a fear that algorithms might reproduce form without a genuine understanding of suffering, beauty, or moral depth.

Despite these concerns, AI-driven automation is widely anticipated to enhance film production efficiency and theoretically expand human creativity. In post-production, AI now automates colour grading, sound mastering, and routine editing—processes that traditionally consumed a substantial amount of time (Izani et al., 2024). Through this shift, filmmakers can ostensibly redirect their focus toward narrative design and aesthetic experimentation rather than mechanical execution.

This technological advantage remains inextricably linked to growing apprehensions regarding diminished originality, ethical ambiguity, data privacy, algorithmic bias, and ownership disputes. Recent scholarship corroborates that while AI integration alters workflows, it simultaneously introduces novel moral and practical challenges that fundamentally redefine authorship and artistic agency (Xu, 2025).

While existing research predominantly emphasises AI's instrumental value and industrial efficiency, this study adopts a divergent perspective. It focuses on two specific phenomena—AI hallucination and AI refusal—frequently dismissed as mere technical flaws, arguing that they carry aesthetic and moral implications in cinematic contexts. By shifting the critical lens from accuracy and productivity to interpretation and affect, this study conceptualises these algorithmic "errors" as generative creative openings rather than systemic malfunctions.

Much of the current literature remains largely descriptive, offering broad overviews of AI applications rather than critically analysing their aesthetic consequences. As Manikandan et al. (2024) note, "the literature currently in publication often provides general summaries of artificial intelligence applications in the film industry rather than going into great depth on the specifics of how AI algorithms impact and improve the visual aesthetics of motion pictures." This observation underscores a critical gap. Despite the burgeoning fascination with AI's creative potential, its substantive influence on cinematic form and ethics remains underexamined.

Methodological Approach and Scope of the Study

To address this limitation, this research employs two complementary methodologies. First, it analyses selected one-minute student short films from Korea university of Media Arts as illustrative case studies. Produced within a pedagogical setting, these works serve as exploratory examples to ground the theoretical discussion on how emerging filmmakers intuitively negotiate automation, creativity, and ethics.

Additionally, it reinterprets AI hallucinations and refusal through an aesthetic-philosophical framework, rather than a strictly technical or regulatory one. Hallucinations, typically categorised as accuracy failures, are reframed herein as a generative deviation stimulating creative insight. Refusal, commonly treated merely as an ethical safeguard, is

reconsidered as an expressive pause facilitating moral reflection and narrative ambiguity.

To contextualise this inquiry, it is imperative to distinguish between rigid moral compliance and broader ethical reflection. Williams (1985) articulates this division: "It peculiarly emphasizes certain ethical notions rather than others, developing in particular a special notion of obligation, and it has some peculiar presuppositions. [...] From now on, therefore, I shall for the most part use 'ethical' as the broad term to stand for what this subject is certainly about, and 'moral' and 'morality' for the narrower system, the peculiarities of which will concern us later on."

Viewed through this theoretical lens, this study strictly delineates "ethical" from "moral." Here, "ethical" refers to the external systemic safety protocols encoded within the AI. Conversely, "moral" denotes the human creator's subjective conscience and internal obligations. The research investigates how human creators exercise moral agency within the ethical boundaries established by technology corporations. Avoiding strictly utopian or dystopian framings, this study seeks to unpack AI's complex implications for contemporary cinema.

This study defines "automation" to resolve conceptual ambiguity. Automation does not refer to the autonomous production of a final cinematic work. It refers specifically to the algorithmic generation of raw visual and narrative materials. This generative process strictly requires human curation. The machine produces unpredictable probabilistic outputs. The human creator then selects, omits, and synthesises these materials. Consequently, automation does not replace human intentionality but functions instead as one interactive component within a continuous aesthetic negotiation.

The primary objective is to propose a conceptual framework exploring how the interplay between algorithmic automation and human curation shapes creative workflows and aesthetic

choices beyond the limiting paradigm of efficiency. While acknowledging that the corpus comprises a limited number of short films from a singular pedagogical setting, these cases function as an exploratory site. This controlled environment allows for a focused examination of micro-level aesthetic decisions rather than attempting to generalise broad industrial shifts.

The secondary objective is to propose a pedagogical framework for film education that integrates AI not only as a production utility but as a dynamic site for moral and aesthetic inquiry. Through this dual focus, this research aims to situate AI-generated cinema within broader academic debates regarding creative agency, media ethics, and social impact, suggesting how machine errors and algorithmic limitations can prompt creators to reconsider the interpretive boundaries of contemporary artistic creation.

Building upon this foundation, the subsequent chapter re-examines hallucination and refusal as phenomena transcending technical irregularities or safety protocols. Rather than viewing them through a corrective lens, it interprets both as aesthetic and moral gestures, and as moments wherein AI unexpectedly participates in the creative process. By tracing how these phenomena disrupt the established logic of algorithmic control, Chapter 2 demonstrates how what is commonly labelled an "error" may serve as a novel principle of artistic production, shifting the discourse from what AI does to what AI means for the future of cinema, film education, and production.

Rethinking Artificial Intelligence, Hallucination and Refusal

Reconceptualizing Technical Errors as Aesthetic Gestures

As outlined in the introduction, both hallucination and refusal have long been regarded as signs of malfunction within

generative AI. In technical and policy-oriented discourse, hallucination denotes the generation of inaccurate, unverifiable, or entirely fictional information, while refusal is understood as a safeguard ensuring ethical compliance. For AI researchers, hallucination functions as a technical term denoting the generation of false data, whereas for film theorists, the same word carries a rich aesthetic lineage, referring to altered perception, imaginative distortion, and the poetics of the unreal. Within a purely technical framework, both phenomena are treated as deviations from the norm and as issues to be corrected to maintain the stability and predictability of AI behaviour.

However, this reductionist framing confines generative AI to the realm of function and control, neglecting the profound interpretive and affective dimensions that emerge when the generative process intersects with artistic creation. Viewed through the lens of cinema, hallucination and refusal cease to be mechanical errors. Instead, they can be understood as creative deviation and moral pause—two complementary gestures that open up exploratory possibilities for pedagogical discussion, conceptualising how creators negotiate the relationship between imagination and responsibility within AI-assisted workflows.

Previous discussions have predominantly framed generative AI's "hallucinations" as technical flaws or prompt failures. Yet, a critical reframing is beginning to emerge even within the technical discourse. As Jiang et al. (2024) assert in their study: "This paper revisits the hallucinations in LLMs, advocating for a paradigm shift where they are not solely viewed as negative phenomena, but also as potential catalysts for creativity." While their observation is rooted in text generation, it provides a crucial interdisciplinary stepping stone.

In cinematic practice, these unpredictable outputs must similarly be re-evaluated as "creative deviations" that open up new possibilities in the creative process. To elucidate this, it is necessary to look beyond the recent technology-centric

literature and turn to the foundational theories that explore the fundamental mechanisms of creativity. Specifically, Boden's cognitive model of creativity provides a robust theoretical framework for explaining how AI's visual anomalies translate into aesthetic value.

Hallucination as Transformational Creativity

Boden (2004) categorises creativity into three distinct dimensions: combinational, exploratory, and transformational. Within this framework, the creative deviation triggered by AI hallucinations is best understood through the lens of "transformational creativity," which breaks the boundaries of established rules and "conceptual spaces." Boden explains this transformative process: "The supposedly impossible idea can come about only if the creator changes the preexisting style in some way. It must be tweaked, or even radically transformed, so that thoughts are now possible which previously (within the untransformed space) were literally inconceivable" (Boden, 2004, p.6).

From this perspective, the AI hallucinations encountered in the filmmaking process can operate as an aesthetic gesture that disrupts and renegotiates the "untransformed space" of the creator's rigid cinematic grammar and representational conventions. Although the visual anomalies generated by AI escape the creator's original intended control, it is precisely at this point of deviation where the filmmaker is confronted with new aesthetic options that were previously unimaginable.

This reinterpretation resonates strongly with cinematic creation. Filmmakers often engage in processes of trial, error, and reinterpretation where unpredictability yields originality. The creative value of such deviation is well articulated by Csikszentmihalyi (1990), who notes that unlike conventional artists who rigidly adhere to a predetermined plan, highly original creators embrace the unpredictability of the process. They typically begin with an intuitive but undefined goal, continuously modifying their work in response to unexpected

elements that emerge. In the context of AI-assisted filmmaking, algorithmic hallucinations provide these unexpected elements, transforming the machine's misinterpretations into spontaneous prompts for aesthetic modification. Seen from this perspective, hallucination transforms from a malfunction to a method—an opening through which artificial systems participate in human creative traditions.

Algorithmic Refusal and the Cinematic Ellipsis

If hallucination represents the expansion of imagination, refusal marks its counterpart—the moment when creation halts in response to encoded boundaries. In the field of AI safety, refusal refers to a model's capacity to decline outputs that violate the societal or legal norms dictated by its algorithmic guidelines (Arditi et al., 2024). Hildebrandt et al. (2025) further conceptualise it as a nonlinear and multidimensional behaviour distributed across the model's layers.

It is crucial to distinguish here between the ethics governing the machine and the morals of the creator. Generative AI does not possess personal subjective morals; rather, it executes a refusal based on externally derived ethical prescriptions. However, in artistic and cinematic contexts, this algorithmic refusal presents a unique situation as it compels the human creator to pause and make aesthetic choices that are underpinned by, and may actively alter, their own subjective matrix of moral beliefs and values. When an AI system declines to visualise violence or harm, it enacts an algorithmic version of silence that invites the human creator to deeply reconsider what should or should not be represented.

This phenomenon parallels the well-established concept of cinematic ellipsis. Throughout film history, moments of deliberate omission—off-screen violence, implied trauma, temporal leaps, and intentional ambiguity—have served as powerful sites of meaning where the unseen resonates more

profoundly than the explicitly visible. Similarly, AI refusal functions as a modern cinematic ellipsis, as an “ethical pause” where the system's alignment protocols intersect with the filmmaker's moral intent. Within this intersection, creativity becomes dialogic. The machine's restriction forces the human artist to search for new symbolic or metaphorical expressions that honour both emotional truth and moral care.

From this standpoint, refusal emerges as more than a technical safeguard. It becomes a pedagogical and reflective moment—a chance for artists to confront the morality of image-making itself. In reframing refusal as an expressive act, AI-driven cinema redefines what it means to “create responsibly.” Silence becomes not the end of expression but its moral beginning.

Taken together, hallucination and refusal embody a dialectic between excess and restraint, imagination and conscience. Hallucination expands the horizon of creative thought through deviation and uncertainty; refusal anchors that expand within moral reflection. As Manovich (2018) observes, AI “plays a crucial role in the global cultural ecosystem” and “helps us make aesthetic decisions when we create media.” Reconsidered as a pair within this culturally embedded system, these behaviours suggest a conceptual reframing of generative AI, positioning it not just as a tool of automation but as a potential medium of interpretation that intersects with the human capacity for creative risk and moral awareness.

These insights provide the conceptual foundation for the next chapter, which examines how hallucinations and refusal materialise within actual AI-generated cinematic works, revealing how generative processes, imagination, and human morality converge in the visual language of contemporary film.

Aesthetic Implications of AI Hallucinations in the Case Studies

Production Parameters and the Analytical Procedure

To address the methodological specificities of AI-assisted cinematic production, Table 1 details the generative environments for the selected case studies. The three specific student films (*Memory in a Pot*, *The Returning Spring*, and *Cut It Off*) were purposively selected from a broader pedagogical corpus because they distinctly isolate the two phenomena under investigation: the first two provide salient visual evidence of an algorithmic hallucination altering spatial and anatomical rendering, while the third offers a documented instance of systemic refusal interrupting the production workflow.

Documenting these parameters is essential for analysing how visual anomalies and systemic refusals manifest, suggesting that these phenomena reflect probabilistic tendencies within the models’ generative processes during iterative workflows, rather than being solely attributable to underspecified prompting.

Crucially, the authors acknowledge a technical inconsistency within the corpus regarding the use of free versus paid model tiers (as detailed in Table 1). Because the projects used both free and paid tiers, this study does not treat hallucination frequency, refusal intensity, rendering quality, or generation time as directly comparable across the cases. Differences in access tier may have affected model behaviour, safety filtering, speed, and output stability. Table 1 is used to contextualise each production workflow rather than support cross-case

Project Title	AI Models & Versions (Tier)	Data Collection Source	Example of Initial Prompt	Generation Iterations & Time
Memory in a Pot	- ChatGPT4 (Free) - DALL-E 3(Free) - Runway Gen-3&Veo2(Free)	Prompt logs, observation, and in-depth interview	In the darkness, a landscape dimly takes shape. Under a cloudy sky, the field sways in the wind. The camera pans slowly, settling on a rural cottage.	Image: approx. 7 attempts Video: 3 iterations - Total generation time: 2 hours
The Returning Spring	ChatGPT 4 (Paid) - DALL-E 3(Paid) Veo2 (Paid)	Interview and prompt logs	Three people sitting at a dining table in complete silence. Jeong-ok quietly ladling out Korean stew and handing it over. Sang-hyun and Hyun-joo accepting the food and starting to eat their meal wordlessly. Cinematic lighting.	- Image: approx. 20 attempts - Total generation time: 3 days
Cut It Off	- ChatGPT4 (Free) - DALL-E 3(Free) - Runway Gen-3 (Free)	Prompt logs, observation, and in-depth interview	Silhouetted parents merely pretending to strike a child	Image: approx. 30 attempts (due to refusals) Video: 5 iterations - Total generation time: 4 days

Table 1 Overview of the AI Generation Parameters and Data Collection
Note: The prompt examples provided in this table represent the initial textual inputs formulated by the students, derived directly from their original screenplays to generate the base images. For conciseness, routine functional commands (e.g., “Generate an image of...”) and the broader conversational contexts within the LLM sessions have been omitted.

causal claims. The analytical procedure relies on a sequential examination of the prompt logs, as well as direct observation and in-depth semi-structured interviews.

First, the prompt logs were reviewed to establish the creator's initial textual intent and to track the deviations generated by the models across multiple iterations. Textual analysis in this study refers to a close reading of the relation between prompts, screenplay intentions, and final AI-generated images. The procedure followed three steps: first, identifying points of divergence between the student's initial screenplay or prompt intention and the generated output; second, describing the specific visual anomalies involved, such as spatial distortion, anatomical instability, atmospheric deviation, or enforced omission; and third, interpreting only the anomalies that were retained or recontextualised by the student creator in the final film. In this sense, the analysis does not measure audience reception or production effectiveness but instead examines how selected creators have assigned aesthetic or moral meaning to algorithmic deviation and refusal.

To limit interpretive overreach, the close readings were checked against the students' prompt logs and interview comments. Eder's (2010) four-dimensional framework was used heuristically to organise the analysis of the AI-generated figure as an audiovisual artifact, fictional entity, symbolic

form, and socio-cultural reference point. This procedure supports the situated interpretation of the selected cases but does not eliminate researcher subjectivity or establish generalisable validity.

Finally, the interview material was utilised directly within the analysis to reconstruct the creator's cognitive processes during moments of technological friction. The interviews were used as contextual materials rather than as a stand-alone empirical dataset. They focused on how the student creators described prompt adaptation, workflow friction, and moral negotiation during production. Because the interview material is cited selectively to clarify specific creative decisions, this study does not claim to offer a systematic qualitative interview analysis or generalisable findings based on coded interview data.

As the previous chapter redefined hallucination as a creative deviation and refusal as moral pause, the present chapter examines how these concepts manifest within actual AI-generated cinematic practice. It focuses particularly on hallucination as a visual and emotional phenomenon that prompts localised, micro-level aesthetic negotiations regarding realism, empathy, and authorship. Through a detailed textual analysis of two student-produced short films, this chapter explores how algorithmic misalignment can produce new aesthetic forms and social meanings.



Fig. 1 Algorithmic hallucination and spatial distortion in *Memory in a Pot* (2025). Source: student production archive.



Fig. 2 Repurposed visual metaphor of *Fading Memory*. Source: student production archive.

Spatial Distortion and Representational Fragility in Memory in a Pot

The first case, *Memory in a Pot* (2025), was produced through an AI-assisted workflow across pre-production, production, and post-production, while remaining dependent on the creator's selective curation. Screenplay elements, imagery, sound, and editing rhythms were generated or refined using the LLMs and video synthesis tools detailed in Table 1. Rather than demonstrating the general capacity of generative AI to act as an autonomous aesthetic collaborator, this case illustrates how one student creator incorporated AI-generated materials into a localised process of selection, omission, and reinterpretation.

A notable instance of AI hallucination appears in the depiction of the elderly mother figure (see Figures 1 and 2). During the image synthesis process, the model produced distortions in spatial depth, bodily proportion, and surface texture. The mother appears in an awkward, partially suspended pose, while parts of her clothing seem to merge with the surrounding background.

Her hands are imprecisely articulated, and the lighting gives the figure a detached, liminal quality within the domestic scene. In the final work, the creator did not simply correct these distortions. Instead, the image was retained and recontextualised as a visual sign of fading memory and emotional fragility. In this limited sense, the algorithmic error became available as material for aesthetic interpretation rather than remaining only a technical defect.

Eder's (2010) framework of character comprehension is used here in a limited and heuristic manner. Since the framework was originally developed for human actors and fictional characters in traditional narrative film, its application to AI-generated synthetic figures should not be understood as a direct methodological transfer. Rather, it provides a provisional structure for organising the close reading of

how the creator assigned meaning to the distorted figure within this specific case.

To limit interpretive overreach, the visual reading focused on the structural anomalies that diverged from the filmmaker's initial narrative intention and were subsequently retained or reinterpreted in the final film. The researcher's interpretation was checked against the student's prompt logs and interview comments, not to establish definitive validity but to clarify how the creator understood and used the generated distortion. This procedure supports a situated reading of the selected case, while acknowledging that the interpretation remains exploratory and case-specific.

In the first dimension, the figure can be read as an audiovisual artifact marked by a distorted posture, unstable texture, and incomplete bodily articulation. These features indicate the model's probabilistic reconstruction of the human form. In the creator's final arrangement, what might otherwise appear as a technical glitch is treated as a visual trace of unstable perception, rather than as evidence of the machine's empathic or intentional vision.

In the second dimension, the figure operates as a fictional entity through which the creator associated the mother with emotional ambivalence, domestic care, and fading presence. The blurred boundary between the body and background can be interpreted, within this film's narrative context, as a visual metaphor for a mother whose presence is emotionally central yet visually unstable. In the third dimension, details such as the shortened grey hair and subdued posture may function as symbolic cues of age, restraint, and memory. These readings are not presented as universal cultural meanings but as context-specific interpretations supported by the film's narrative and creator's account.

Finally, in the fourth dimension, the image may be situated within contemporary digital visual culture where

synthetic images often aspire to smoothness, coherence, and a hyper-realistic finish. Against this expectation, the distorted mother figure draws attention to the instability of AI-generated representation. This does not establish how audiences receive the image nor does it prove that algorithmic imperfection inherently produces emotional authenticity. It shows more modestly how the creator recontextualised a generated anomaly as an interpretive metaphor within this particular pedagogical case.

The student director of *Memory in a Pot* stated in the interview: “It was only when human sensibility and direction were added to the AI’s creations that the warmth of the work could finally be felt.” This comment is used here as contextual evidence of the creator’s own understanding of the workflow. It suggests that, in this case, the meaning of the hallucinated image emerged not from the algorithmic output alone but from the creator’s subsequent decisions to retain, frame, and reinterpret the distortion within the film’s narrative.

Atmospheric Deviation and Transformational Curation in The Returning Spring

The second case, *The Returning Spring* (2025) offers an ideal site for observing how algorithmic unpredictability stimulates creative learning and stylistic transformation (see Figures 3



Fig. 3 Curated atmospheric deviations in *The Returning Spring* (2025). Source: student production archive.

and 4). During production, the student director repeatedly attempted to generate a specific realistic cinematic space. However, despite over 20 prompt iterations (as shown in Table 1), the system consistently misread depth, lighting, and atmosphere, introducing thick fog, heavy shadows, and spatial distortions absent from the original storyboard.

Rather than abandoning the process due to these “errors,” the filmmaker began to collaborate with the machine’s unpredictability. He achieved the final cinematic space through the following process. First, he provided a reference image that closely resembled his original vision. Then, as shown in Figure 4, he specified the camera angles and framing to generate an output as close to his desired imagery as possible. Although he ultimately failed to precisely render the intended spatial combination of a vegetable garden alongside the grandmother’s house, he successfully generated a satisfactory atmospheric image, as seen in Figure 3, and integrated it as the primary setting for his film.

This process can be read through Boden’s concept of transformational creativity, insofar as the student’s initial storyboard was modified in response to the AI’s atmospheric deviations. Rather than demonstrating a general mechanism of creative transformation, the case illustrates a localised

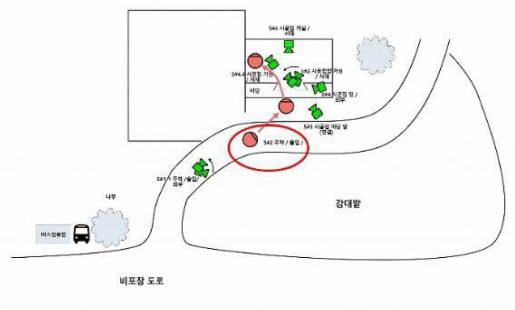


Fig. 4 Initial reference image for aesthetic negotiation. Source: student production archive.

instance in which an unintended output was selectively retained and reinterpreted as part of the film's visual language.

As Csikszentmihalyi (1990) suggests, successfully navigating unexpected creative challenges induces a state of optimal experience. By continuously negotiating with these unpredictable hallucinations, the filmmaker integrated the fog and shadows to reflect the protagonist's psychological isolation, elevating the aesthetic complexity of the work. This exemplifies what Jiang et al. (2024) describe as the divergent phase of creative reasoning, in which deviation fosters discovery. The case suggests that negotiating between intention and contingency encouraged the creator to revise the original visual plan and to assign new narrative value to the unexpected fog and shadows. However, the analysis remains limited to the documented production process and does not claim to measure broader psychological or educational development.

Within the scope of these pedagogical experiments, this dynamic suggests potential aesthetic shifts that merit further exploration in the context of digital visual culture. As contemporary media becomes increasingly saturated with hyper-realistic, flawlessly polished synthetic images, these student works suggest an exploratory alternative for audience engagement. Through the aesthetic lens of these case studies, authenticity can be conceptually framed as a negotiation centred on vulnerability and imperfection rather than technical perfection. In this context, the structural anomalies and 'glitches' of AI hallucination are interpreted by the creator as textual markers of fragility, resisting the sterile predictability of standard algorithmic outputs. Imperfection, once a sign of machine limitation, is repurposed as a marker of human-like fragility and empathy.

Hallucination should not be treated merely as a defect to be corrected but as a pedagogical condition; a state of productive uncertainty that cultivates reflection, interpretation, and critical awareness. Through this reframing, the pedagogical space

of AI-assisted filmmaking serves as an exploratory site where technology and humanity meet, not in total control or mechanical replication, but in dialogue and mutual understanding.

Refusal in AI Cinema and Film Education

The Shift from Algorithmic Safety to Aesthetic Restraint

Building upon the previous chapter's discussion of hallucination as creative deviation, this chapter examines the phenomenon of refusal—moments when generative AI declines to produce certain outputs or representations. While hallucination reveals the productive potential of error, refusal exposes the aesthetic and pedagogical value of limitation. In both instances, creativity emerges not from absolute control but from negotiation, the dynamic interplay between what the machine generates and what it withholds.

As established in Chapter 2, algorithmic refusal functions technically as a safety mechanism that filters prompts violating encoded societal norms (Arditi et al., 2024; Hildebrandt et al., 2025). However, when transported from the domain of AI safety into the realm of artistic production, the function of this mechanism fundamentally shifts. It is vital to distinguish between the machine's ethics and the creator's morals. Generative AI does not possess personal, subjective morals; it operates strictly on encoded ethical prescriptions. From an artistic perspective, this algorithmic refusal transcends passive prohibition. It introduces moments of silence and restraint that actively compel human users to pause and reflect upon their own internal, subjective moral boundaries regarding representation.

Examined through the lens of cinematic practice, such refusals acquire profound aesthetic and methodological significance. They mirror the intentional ellipses long present in cinematic history—moments when filmmakers deliberately withheld explicit images of violence or suffering. Within AI-mediated

creation, the system's ethical restriction thus functions as a contemporary continuation of this cinematic omission.

Negotiating Systemic Limitations in Cut It Off

A clear example of algorithmic refusal acting as a catalyst for creative negotiation appears in *Cut It Off* (2025), an AI-generated student short film addressing the eradication of child abuse. The filmmaker initially sought to explicitly visualise domestic violence to raise awareness of its intergenerational cycle. However, the generative systems consistently refused to depict scenes involving physical harm to children due to their safety alignment protocols.

As the system continued to restrict violent imagery, the filmmaker explored alternative prompt strategies to symbolise the trauma. For instance, the student input prompts requesting "silhouetted parents merely pretending to strike a child," attempting to bypass the explicit violence filter. Yet the system still triggered strict safety protocols against implied abuse, halting the generation process entirely.

To overcome this persistent algorithmic refusal, the student adopted a fragmented, hybrid production workflow. The filmmaker generated the images of the parents and child



Fig. 5 Isolated generation of components to circumvent algorithmic refusal in *Cut It Off* (2025). Source: student production archive.

separately as isolated "safe" components. These individual elements were then manually composited into a single frame using Adobe Photoshop. Finally, this composited image was animated using the image-to-video function of Runway Gen-3 (Free tier).

The Aesthetics of Invisibility and Moral Reflection

What began as a technical obstacle evolved into a moment of methodological and moral reorientations. By forcing the filmmaker to omit direct depictions of violence, the AI's refusal accidentally enforced a moral distance. This aesthetic of invisibility deeply resonates with Sontag's (2003) critique of visual culture, which warns that explicitly exposing the pain of others often reduces their suffering to a consumable spectacle, leading to moral anesthetisation. The resulting cinematic ellipsis prevents the trauma from being objectified as a spectacle.

As Rancière (2009) argues in his concept of the "emancipated spectator," the omission creates a rupture—a blank space



Fig. 6 Original idea image prior to algorithmic refusal. Source: student production archive.

where the audience is no longer a passive consumer of explicit images but an active participant who must morally and intellectually reconstruct the meaning of the unseen.

In the case of *Cut It Off*, the filmmaker's repeated engagement with algorithmic refusal can be interpreted as generating three case-specific learning points rather than measured pedagogical outcomes:

1. Moral Awareness: The student recognised that artistic freedom is inseparable from subjective moral responsibility.
2. Symbolic Thinking: Forced to work within machine-imposed ethical limits, she learned to communicate meaning through metaphor and aesthetic omission rather than literal depiction.
3. Empathic Perspective: The shift from showing violence to evoking empathy redefined the purpose of representation from mere exposure to deep understanding.

Within the cinematic tradition, silence, absence, and suggestion have long served as powerful aesthetic devices. Similarly, AI's refusal introduces an aesthetic of invisibility—a method of generating meaning through what remains unseen. When a model declines to visualise exploitation, it implicitly invites the filmmaker to consider the moral weight of visibility itself.

This dynamic can be productively recontextualised through Levinas' philosophy, particularly his notion that morality begins with the encounter with the "face" of the Other, an obligation that precedes any conceptual representation (Levinas, 1985). Crucially, because the machine itself lacks a subjective conscience, its algorithmic refusal must not be analogised to an ethical subject. The system's structural refusal presents a rigid, unyielding 'mechanical wall' that abruptly interrupts the creative workflow. Rather than acting as a moral agent, this digital friction serves as a passive catalyst that forces the human creator to look inward and

confront their own internal obligations toward the unrepresented subject. Within this framework, Levinas's assertion that the face says "thou shalt not kill" (Levinas, 1985, p. 87) is enacted not by the algorithm but within the human conscience, which is forcefully reactivated precisely at the moment of technological disruption, thereby repositioning the filmmaker as the primary, active agent of moral care within the boundaries of image-making.

This understanding reframes the role of technology within these specific pedagogical contexts. Generative AI remains a non-subjective technical system, but its refusals can function as points of friction through which creators examine their own representational responsibilities. The system does not conduct moral reasoning; rather, its constraints redirect moral reasoning back to the human filmmaker. In doing so, it prompts an exploratory approach to algorithmic creativity not as the unlimited production of images but as the thoughtful discernment of representational boundaries and selective omission over the unlimited production of images.

Generative AI's refusal reveals that limitation and creativity are interdependent forces. Just as hallucination expands the cinematic imagination through unpredictable deviation, refusal deepens it through moral reflection. Together, they form a powerful aesthetic-moral dialectic in contemporary filmmaking—a dynamic wherein algorithmic excess and ethical restraint, visual invention and cinematic silence, coexist in productive tension.

Rather than presenting AI as a moral agent, this dynamic positions algorithmic constraint as a methodological interruption that can be incorporated into cinematic language through human interpretation and curation. These insights prepare the ground for the final chapter, which synthesises how hallucination and refusal move beyond technical categorisations to offer a pedagogical and methodological framework for aesthetic-moral gestures in digital film production.

Conclusion

The Aesthetic-Moral Dialectic in Digital Filmmaking

The integration of generative artificial intelligence into cinematic practice necessitates a critical re-evaluation of digital aesthetics and creative agency. Moving beyond the dominant discourse of technical efficiency and unidirectional automation, this study has synthesised two phenomena commonly dismissed as algorithmic flaws—hallucination and refusal—to propose them as foundational elements for a critical pedagogical framework in film education and emerging production workflows. Rather than merely summarising these behaviours as isolated technical events, this study proposes that they can be understood conceptually as interdependent aesthetic-moral gestures. Through the dialectic of deviation and restraint, the interplay between human curation and algorithmic constraints offers a novel framework for exploring contemporary digital authorship.

Within the observed workflows, generative AI functions not primarily to perfect photographic accuracy, but to catalyse new interpretive practices. Specifically, it compels creators to reinterpret algorithmic anomalies and systemic limitations as functional visual metaphors. As demonstrated through the case studies, hallucination operates as a mechanism of transformational creativity. It exposes the interpretive gap between algorithmic probability and human perception, turning technical misalignment into a generative force that visualises the unseen. Conversely, refusal functions as a digital cinematic ellipsis, as an ethical interruption that demands the creator pause and reflect on the moral weight of representation. Together, these forces illustrate that algorithmic art and human morality are not opposing domains, but intertwined processes of aesthetic negotiation.

The broader implication of this aesthetic-moral dialectic should therefore be understood as conceptual rather than empirically verified. Within the selected student works,

structural anomalies of AI hallucination were not treated solely as technical deficiencies; they were recontextualised by the creators as signs of fragility, memory, or uncertainty. This suggests that imperfection may function as an aesthetic resource in AI-assisted filmmaking, although further audience research would be required to determine how viewers perceive such anomalies.

Pedagogical Implications for Future Cinematic Practice

While this critical reframing shift fundamentally alters cinematic aesthetics, it simultaneously opens up vital new possibilities within film education. Traditional film pedagogy has often centred on technical mastery and the exertion of absolute control over the frame. Yet, encountering the unpredictability of hallucination and the strict boundaries of refusal invites a necessary shift from pure production to profound reflection. By engaging with the machine's deviations, students cultivate critical literacy and symbolic thinking; they learn to read ambiguity as insight and silence as expression. Film education must embrace these algorithmic behaviours not as obstacles to overcome, but as a method that turns machine error into human inquiry and algorithmic limitation into moral creativity.

This study suggests that the convergence of hallucination and refusal indicates emerging modes of moral and aesthetic negotiation in cinematic creation. Generative AI challenges the historical dominance of the director's absolute precision, reminding us that profound aesthetic beauty often arises from relinquishing control to uncertainty. Simultaneously, its algorithmic refusals remind us that silence carries immense moral depth—that the conscious decision not to visualise is itself a powerful cinematic gesture.

The cases examined here suggests that AI-assisted cinematic practice may benefit from neither rejecting automation outright nor treating it as a neutral efficiency tool. Instead, these cases show how selected creators have negotiated

with algorithmic outputs and constraints through acts of selection, omission, and reinterpretation. In this limited sense, hallucination and refusal can be understood as useful concepts for teaching students to reflect on the aesthetic and moral boundaries of digital image-making.

More concretely, this framework can be organised into three teaching stages. The first stage, prompt and anomaly documentation, asks students to record prompts, outputs, refusals, and unintended deviations. The second stage, aesthetic reinterpretation, requires students to select one hallucination or refusal and justify its narrative or ethical function through close visual analysis. The third stage, reflexive revision, asks students to produce a short AI-assisted sequence and submit a process log explaining how algorithmic constraints shaped their representational choices. The learning objectives are to develop critical AI literacy, symbolic visual thinking, and reflective moral judgment. Assessment should focus on the clarity of the prompt documentation, the coherence of aesthetic justification, and the student's ability to explain the ethical implications of omission, restraint, and curation.

References

- Arakawa, K. (Director). (2016). *Never-ending man: Hayao Miyazaki* [Documentary]. NHK.
- Arditi, A., Obeso, O., Syed, A., Paleka, D., Panickssery, N., Gurnee, W., & Nanda, N. (2024). Refusal in language models is mediated by a single direction. *Advances in Neural Information Processing Systems*, 37, 136037-136083.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
- da Silva Oliveira, M. S., de Moraes Sato, P., Arceno, M. A., Ulian, M. D., Unsain, R. F., Cardoso, M. A., & Scagliusi, F. B. (2022). "It's women's obligation:" constitutive elements of gendered domestic cooking practices performed by women from western Brazilian Amazon. *Food, Culture & Society*, 25(3), 540-560.
- Eder, J. (2010). Understanding characters. *Projections*, 4(1), 16-40.
- Hildebrandt, F., Maier, A., Krauss, P., & Schilling, A. (2025, June). Refusal behavior in large language models: A nonlinear perspective. In *2025 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.
- Hinsz, V. B., Matz, D. C., & Patience, R. A. (2001). Does women's hair signal reproductive potential? *Journal of Experimental Social Psychology*, 37(2), 166-172.
- Izani, M., Razak, A., Faiz, M., Kaleel, A., & Assad, A. (2024, July). Harmonizing control and creativity with AI for animator's toolkit. In *2024 IEEE Symposium on Industrial Electronics & Applications (ISIEA)* (pp. 1-7).
- Jiang, X., Tian, Y., Hua, F., Xu, C., Wang, Y., & Guo, J. (2024). *A survey on large language model hallucination via a creativity perspective*. arXiv preprint arXiv:2402.06647.
- Levinas, E. (1985). *Totality and Infinity: An Essay on Exteriority*. Duquesne University Press.
- Manikandan, C., Kashyap, A., & Nahak, F. M. (2024, February). Discourse of AI-Influence in Visual Aesthetics. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, 5, pp. 1127-1130).
- Manovich, L. (2018). *AI aesthetics*. Strelka Press.
- Rancière, J. (2009). *The Emancipated Spectator* (G. Elliott, Trans.). Verso.
- Sontag, S. (2003). *Regarding the Pain of Others*. Farrar, Straus and Giroux.
- Williams, B. (1985). *Ethics and the limits of philosophy*. Harvard University Press.
- Xu, Y. (2025). *Balancing Creativity and Automation: The Influence of AI on Modern Film Production and Dissemination*. arXiv preprint arXiv:2504.19275.