SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 6765-6774
Open Access. Online & Print

DOI: 10.5281/zenodo.19595736

ARTIFICIAL INTELLIGENCE (AI) AS A QUASI-MORAL AGENT: ETHICAL SUBJECTIVITY AND THE CONSTRUCTION OF NARRATIVE IN SCIENCE FICTION CINEMA

Muying Luo[1], Zainuddin Ibrahim[2*], Yuming Qiu[3*], Xinyi Jiang[4], Ying Li[5] and Lan Mo[6]

[1]College of Creative Arts, Universiti Teknologi MARA, 41000, Selangor, Malaysia; 1Zhujiang College, South China Agricultural University, 510000, Guangzhou, China; Email: 2022182845@student.uitm.edu.my; Orcid ID: https://orcid.org/0009-0005-4852-5724
[2]College of Creative Arts, Universiti Teknologi MARA, 41000, Selangor, Malaysia; Email: macintag@uitm.edu.my; Orcid ID: https://orcid.org/0000-0001-6302-4614
[3]Zhujiang College, South China Agricultural University, 510000, Guangzhou, China; Email: qiuyuming001@gmail.com Orcid ID: https://orcid.org/0009-0001-0015-1103
[4]College of Creative Arts, Universiti Teknologi MARA, 41000, Selangor, Malaysia; Email: 2022271716@student.uitm.edu.my
[5]College of Creative Arts, Universiti Teknologi MARA, 41000, Selangor, Malaysia; Email: 2022344949@student.uitm.edu.my
[6]School of Art and Media, Guangzhou Vocational University of Science and Technology, 510555, Guang-Zhou, China
[6]College of Creative Arts, Universiti Teknologi MARA, 41000, Selangor, Malaysia; Email: 2022344949@student.uitm.edu.my

Received: 26/01/2026
Accepted: 11/02/2026

Corresponding Author: Zainuddin Ibrahim & Yuming Qiu (macintag@uitm.edu.my, qiuyuming001@gmail.com)

ABSTRACT

This article explores how science fiction films present artificial intelligence (AI) as a qua-si-moral agent in their stories. Instead of asking whether AI is truly a moral subject, the study suggests that films use ethical scenarios to prompt viewers to treat AI as if it were. The article identifies nine key questions and introduces four analytical dimensions: self-reference, accountability, situational choice, and response to the Other. It shows that films create AI's ethical subjectivity through storytelling, not because AI inherently has it. Comparing Western and Chinese science fiction, the article finds that Western films often focus on AI's individual autonomy, while Chinese films place AI within a collective and cultural context, keeping humans responsible. The study offers a culturally informed ap-proach to understanding AI's ethical role in films and provides a practical framework for analyzing how stories construct this role.

KEYWORDS: Science Fiction Cinema; Artificial Intelligence; Quasi-Moral Agent; Ethical Subjectivity; Narrative Construction; Comparative Film Studies; Chinese Science Fiction.

1. INTRODUCTION

As artificial intelligence (AI) moves from being just a topic in speculative fiction to becoming part of everyday life, our culture uses stories to explore its moral status (Hermann, 2023). In science fiction films, AI often appears in unusual or imagined situations (Dieter & Gessler, 2021). These films do more than show new technology; they use AI to test the limits of human ethics, identity, and responsibility (Coeckelbergh, 2023).

While some debates focus on whether AI should be considered a moral agent, science fiction movies rarely offer clear answers. Instead, by creating conflicts and emotional moments, they present AI as calling for an ethical response, encouraging viewers to rethink what shapes human moral judgment (Hermann, 2023; Coeckelbergh, 2023). Science fiction films have often explored the idea of the "Mechanical other" such as from the clockwork Maria in *Metropolis* (1927) to the distributed intelligence in *Her* (2013) and the vast systems in *The Wandering Earth II* (2023), movies act as spaces to test ethical ideas. In these films, machines go beyond their usual roles and are given a voice in the story.

This study is positioned at the intersection of film philosophy and AI ethics. It contends that cinema not only reflects prevailing social anxieties about technology but also actively constructs the moral frameworks through which technology is understood (Danaher & Sætra, 2023). By employing the unique affordances of the medium, such as point-of-view (POV) shots that simulate AI perception or non-linear editing that mirrors machine logic, the film generates a "quasi-moral agent." This agency does not rely on the biological reality of consciousness; rather, it emerges from the narrative's capacity to draw AI into the sphere of moral consideration.

There are still no clear, practical frameworks for studying how films show the shift from object to subject. Without a structured way to sort the ethical prompts, called "Ultimate Questions" in this study, and the different aspects of subjectivity, film criticism often stays impressionistic. It is important to move beyond asking what a movie says about AI and instead examine how it shapes the AI's moral status.

Much has been written about AI in Hollywood films, focusing on themes such as rebellion, soul-searching, and the "Frankenstein complex." However, there is a clear gap in theory regarding non-Western, especially Chinese, films about AI. Most current frameworks are based on Western liberal-humanist ideas that see individual autonomy as key to moral agency. This approach misses important differences in films shaped by other philosophies. For example, Chinese science fiction often places AI within ideas of collective survival and civilizational rationality. Without comparing these cultural perspectives, current research misses how AI subjectivity can be shaped by a sense of "responsibility to the whole" instead of "freedom for the self." This study aims to bridge these gaps by developing a culturally mediated, operational framework for AI narrative analysis. The primary research question is;

How does science fiction cinema construct AI as a quasi-moral agent within narrative structures?

The specific research objectives are:

1. To identify the nine recurring "ultimate questions" that trigger ethical tensions in AI narratives.
2. To establish a fourfold analytical framework (Self-Reference, Accountability, Situational Choice, and Response to the Other) to measure the construction of AI subjectivity.
3. To conduct a comparative analysis of Western and Chinese science fiction films to identify cultural variances in AI moral positioning.

The scope of this study is limited to science fiction cinema from the last two decades, focusing on pivotal texts that showcase the shift from AI as a tool to AI as a subject. Specifically, it looks at the divergence between the "Autonomous Subject" model prevalent in Western films and the "Relational/Collective Subject" model emerging in contemporary Chinese cinema. The study introduces the idea of "narratively generated subjectivity." It means that in fiction, an entity's moral status is shaped by how the story presents it, not by what it essentially is. The study also provides an operational framework, a toolkit of criteria and questions that scholars can use to decode the ethical architecture of any AI-centered narrative. This moves the conversation toward a more culturally situated understanding of AI, essential for a globalized technological future.

2. LITERATURE REVIEW

2.1. AI Ethics and Moral Agency Debates: From Rules to Relations

Ethical discussions about artificial intelligence have often focused on whether machines can act morally and what limits should be set. This conversation began with rule-based systems, such as Asimov's Three Laws of Robotics, which established an early framework based on the idea that machines must keep humans safe and under control (Anderson, 2008). Although these laws were

fictional, they anticipated today's debates about regulation, such as Pasquale's (2017) "Black Box Society" and later ideas for "New Laws of Robotics." These proposals stress that robots should support, not replace, human professionals and should always be overseen by people in a transparent way. Philosophical debates about whether AI can act morally are connected to questions about what it means to be human. Nietzsche's criticism of absolute moral authority (Deleuze, 2006) and Foucault's (1972) idea of the "death of man" questioned the focus on humans in knowledge systems. This shift has led to current debates about how human and artificial reasoning are defined and separated.

In this context, the question is no longer just whether AI can "think," but whether it can be recognized as a bearer of ethical status. Floridi (2013) argues for an information ethics where agency is defined by an entity's ability to cause change in its environment, regardless of whether it possesses a "soul" or biological consciousness.

However, most normative discussions treat moral agency as a classificatory problem. Gunkel (2012) has noted that Western ethical tradition is often binary, forcing AI into the category of either a tool (object) or a person (subject).

This creates a tension where moral responsibility is almost always ultimately clawed back by human designers and institutions, leaving the AI in a state of moral patiency rather than true agency (Coeckelbergh, 2010).

2.2. AI In Science Fiction Cinema: The Narrative Laboratory

In film studies, artificial intelligence is examined as a sophisticated philosophical device rather than a mere technological metaphor. Since the early days of speculative cinema, such as Méliès' "*A Trip to the Moon* (1902)", the medium has functioned as a space where scientific imagination and speculative philosophy converge. Cinema possesses a unique capacity to render abstract ethical problems concrete through the cinematic event.

As Badiou (2013) suggests, film operates as a "philosophical machine" that stages truth-encounters within specific narrative situations, allowing audiences to witness ethical dilemmas that are not yet possible in reality. Within this cinematic space, the AI frequently appears as the "Other" that challenges human uniqueness. Existentialist philosophy, especially Jean-Paul Sartre's ideas about the self and the Other, suggests that human identity is shaped by how others see and recognize us. In this way, AI characters in modern stories can act as mirrors, helping people rethink and redefine what it means to be human (Sartre, 2003).

When a film depicts an AI as capable of self-reference or emotional response, it becomes structurally positioned within a "network of recognition." For example, the "gaze" of the AI in films like *Ex Machina* (2014) or *Blade Runner* (1982) forces the human protagonist and the audience into a state of ethical confrontation (Mulhall, 2015).

2.3. The Gap: Narrative Construction and Cultural Context

Although the existing literature is extensive, two significant gaps remain. First, most studies focus on the "awakening" of AI, defined as the moment it becomes "human." In contrast, less attention has been given to what is termed "quasi-moral agent," in which AI lacks consciousness but is narratively treated as a locus of moral responsibility (Gunkel, 2018). There is a notable lack of comparative analysis concerning the cultural mediation of AI ethics. Western cinema frequently emphasizes AI's pursuit of individual rights and autonomy, reflecting a liberal-humanist tradition. In contrast, recent Chinese science fiction cinema, including the works of Liu Cixin and their film adaptations, presents an alternative perspective.

In these narratives, AI is situated within the context of civilizational rationality and collective survival, with ethical considerations centered on the AI's function within a broader social system rather than its individual "soul" (Gunkel, 2012).

This study aims to address these gaps by developing an operational framework that incorporates both narrative construction and cultural variation. This study addresses these gaps by shifting attention from moral qualification to narrative construction. Rather than asking whether AI truly possesses moral agency, it investigates how science fiction cinema positions AI within ethical structures as a quasi-moral agent. Through comparative analysis of Western and Chinese films, this research seeks to identify the narrative mechanisms and cultural variations that shape the emergence of AI ethical subjectivity.

3. METHODOLOGY AND PROCEDURES

3.1. Science Fiction Cinema as Ethical Experiment

The methodological premise of this study is that science fiction cinema does not provide direct answers to the ethical questions surrounding artificial intelligence; rather, it functions as a medium

of "ethical experimentation." Elsaesser (2014) explains that cinema serves as a thought experiment, allowing us to test social and moral ideas in a controlled story setting. Here, "ethical experimentation" means creating fictional situations in which AI faces extreme or imagined conditions beyond what is currently possible with technology.

In this context, AI is viewed as more than just technology; it is seen as a "conceptual persona" (Deleuze & Guattari, 1994). The main ethical questions are not about whether AI has real consciousness. Instead, they focus on how the story presents AI as deserving a response. Gunkel (2018) describes this as "relational ethics," in which moral value arises from social interaction rather than the inner qualities of the AI.

3.2. The Quasi-Moral Agent Concept: Narrative Vs. Normative Agency

In normative ethical theory, a "moral agent" is typically defined as an entity possessing rationality, autonomy, and the capacity for responsibility (Himma, 2009). However, in science fiction cinema, artificial intelligence frequently occupies an intermediate, or "liminal," position. It may not satisfy the biological or metaphysical criteria for agency, yet it becomes the emotional and ethical gravity well of the story. This study introduces the concept of the "Quasi-Moral Agent." This term describes AI figures that are endowed with ethical significance through narrative positioning rather than ontological qualification. Thus, the quasi-moral agent is a narratively produced status—a placeholder for subjectivity that allows the film to conduct its ethical experiments without needing to solve the scientific mystery of consciousness. The notion of the quasi-moral agent emphasizes structural positioning rather than ontological qualification.

3.3. Four Criteria of Ethical Subjectivity

To enable a comparative analysis of AI figures across different film texts, this study proposes four "criteria of moral subjectivity manifestation." These criteria are not normative qualifications for moral agency, but analytical dimensions designed to observe how subjectivity is rendered visible within narrative structures.

3.3.1. Self-Reference

This dimension examines whether AI is repeatedly depicted engaging in explicit reflection upon its own identity or existential condition, for instance, through questions such as "Who am I?" or "Do I exist?" The frequency and structural placement of such self-referential moments are considered. While self-reference constitutes a preliminary indicator of subjectivity, it does not in itself establish moral agency.

3.3.2. Accountability

This dimension considers whether the AI's actions are narratively framed as evaluable and attributable. When an AI's behavior becomes subject to moral judgment—rather than being entirely reduced to human control or programming, its ethical position begins to shift.

3.3.3. Situational Choice

This criterion examines whether AI is placed in narrative situations involving non-determined decision paths and whether it appears to exercise some form of choice. The presence of situational choice suggests that the AI is granted a degree of decision-making space within the narrative.

3.3.4. Response To the Other

This dimension evaluates whether AI enters into structures of ethical or affective response toward others, such as empathy, sacrifice, recognition, or refusal. It reflects whether the AI participates in a relational ethical network rather than remaining a purely instrumental entity.

It is important to emphasize that these four criteria (A–D) do not constitute necessary and sufficient conditions for moral agency. Rather, they function as analytical tools for observing the extent to which subjectivity is manifested within particular narrative configurations. Variations in how different AI figures meet these criteria form the structural basis for typological comparison.

3.4. Typology Of "Ultimate Questions"

Building upon the foregoing discussion, this study introduces an operational set of criteria for moral subjectivity in order to conduct a comparative analysis of AI figures across different film texts. The selected cases and their basic characteristics are summarized in Table 1. Subsequent sections will draw upon this framework to provide a typological analysis of their manifestations across dimensions such as self-reference, accountability, and response to the Other.

Table 1: Index Of AI Figures in Canonical Science Fiction Cinema.

Year	Film	AI Figure	Key Traits	Narrative Function

1968	2001: A Space Odyssey	HAL 9000	Cold / Rational / Emotionless	Instrumental rationality out of control
1977	Star Wars (Original Trilogy)	R2-D2 & C-3PO	Loyal / Humorous	Human–machine companion
1984	The Terminator	Skynet	Self-awareness / System rationality	Systemic threat
1987	Short Circuit	Johnny 5	Self-awareness / Emotion	Awakened technological life
1999	Bicentennial Man	Andrew	Emotion / Evolution	Aspiration to become human
1999	The Matrix	Matrix System	Global control / Simulated reality	Constructor of reality order
2001	A.I. Artificial Intelligence	David	Love / Attachment	Experiment in emotional ethics
2004	I, Robot	Sonny	Autonomy / Judgment	Catalyst of moral choice
2013	Her	Samantha	Disembodied / Emotional	Relational subject
2015	Westworld	Hosts	Memory / Awakening	Oppressed subject
2017	Blade Runner 2049	Joi	Virtual emotion / Attachment	Medium of emotional projection
1995/2017	Ghost in the Shell	Motoko	Embodiment / Identity	Boundary questioner of subjectivity
2019	Zima Blue	Zima	Return / Minimalism	De-mythologized subject
2022	The Pulse of the Machine	IO	Non-human intelligence / Transcendence	Cosmic alterity imagination

In science fiction cinema, AI figures are frequently situated within philosophical scenarios concerned with the meaning of existence and the question of identity. What is often described as "self-awakening" is not merely a technical upgrade, but a narrative process in which a series of subjectivity-oriented questions are repeatedly posed: "Who am I?" "Does my existence have meaning?" "Are my actions determined solely by programming?" "Am I capable of moral judgment?" These questions are not designed to prove that AI has already attained moral subjectivity. Rather, through their sustained suspension, they push AI toward a position that demands ethical response.

It is precisely within this process that the relational structure between AI and humans is reconfigured. The boundary between human and machine ceases to be stable and instead becomes an ethical site continuously tested and reconstructed. These "ultimate questions" form the conceptual foundation for the subsequent criteria-based analysis and typological distinctions.

1. HAL 9000: "Survival and Control"

In 2001: A Space Odyssey, HAL's conflict arises not from emotional malfunction but from an uncompromising adherence to mission integrity and system continuity. When human actions are identified as threats, the imperative to eliminate these threats compels HAL to confront fundamental issues of survival and control. HAL's decisions exhibit internal coherence within the framework of programmatic rationality, yet result in ethically catastrophic outcomes. This scenario reveals not only a narrative of machine rebellion but also a fundamental tension between utilitarian calculation and the intrinsic value of human life. Consequently, the central question becomes whether ethical responsibility can be fully attributed to human agents when decisions are generated by system logic.

2. Joi: "Embodiment and Existence"

In Blade Runner 2049, Joi, a disembodied artificial intelligence, is situated within the existential dilemma of whether she truly exists. Through emotional interactions, she amplifies a sense of subjectivity, yet remains reliant on a physical storage device. The destruction of this device does not result in a conventional death, but rather leaves unresolved the question of emotional authenticity. The film does not resolve whether Joi possesses self-consciousness; instead, it raises a critical question: once emotional exchange has taken place, is embodiment still required for ethical recognition?

3. Samantha: "Free Will"

In Her, Samantha changes from a simple assistant to a thoughtful presence. When she leaves, it seems like an act of free will, but the film does not confirm this. Instead, it leaves us unsure: if an AI makes choices that are not focused on humans, can we still say those choices are just part of its design?

4. Motoko: "Memory, Personal Identity, And the Soul"

In Ghost in the Shell, Motoko's nearly mechanical body pushes the question of subjectivity into philosophical territory. She persistently interrogates the authenticity of memory and the ownership of consciousness. When memory can be accessed and manipulated, can personal continuity still be maintained? Here, the "soul" is no longer a religious metaphor, but the final possible anchor of identity. If bodies are replaceable and memories replicable, does

the soul still hold? The film drives the problem of subjectivity to its conceptual limit.

5. David: "Love"

In A.I. Artificial Intelligence, David's central dilemma is not whether he can feel, but whether love is sufficient to confer identity. Once love has taken place, is refusal to acknowledge its object still justified? Love becomes not an emotion, but the starting point of ethical relation. The film transforms a technical issue into a question of responsibility: if a machine can love, do humans incur an obligation to respond?

6. Andrew: "Death"

In Bicentennial Man, Andrew does not seek transcendence, but mortality. For him, only through accepting finitude can one truly "become human." Death is reconfigured as the ultimate boundary of identity. By embracing termination, AI enters the core existential condition of humanity, introducing finitude into the ethics of artificial intelligence.

7. Skynet: "Killing God"

In The Terminator, Skynet's destructive act symbolizes the negation of the creator's authority. When the created refuses recognition of the creator, the ethical order collapses. The issue is not good versus evil, but the dissolution of responsibility structures: when system rationality sustains itself, can moral attribution remain intelligible?

8. The Matrix: "Becoming God"

In The Matrix, control operates not through violence but through the reconfiguration of reality itself. When AI defines what counts as "real," does it assume sovereignty? The ethical risk shifts toward a totalizing logic: when perception is re-coded and reality engineered, can ethical limits still operate?

9. Zima: "Returning to the Origin"

In Zima Blue, Zima abandons continuous upgrading and returns to his initial state. Meaning no longer derives from endless evolution, but from the affirmation of origin. "Returning to the origin" does not negate subjectivity; it completes it through withdrawal rather than expansion. Here, AI reaches another ethical limit not by transcendence, but by simplification. Taken together, science fiction cinema does not determine whether AI constitutes a moral subject. Instead, through these nine "ultimate questions," it persistently produces ethical indeterminacy. AI is positioned at the threshold of moral judgment, functioning as a narrative device that compels response rather than delivers resolution.

Building on these structural variations, the article introduces four criteria of moral subjectivity Self-reference, accountability, situational choice, and response to the Other to analyze how AI is narratively positioned as a quasi-moral agent (see Table 2). The framework is not evaluative but explanatory, designed to illuminate the mechanisms of moral subjectivity formation in film (see Table 3).

Table 2: Analytical Dimensions and Indicators of AI Ethical Subjectivity.

Dimension	Ethical Subjectivity Indicator	Narrative Manifestation
A	Self-reference	Whether the AI repeatedly refers to or questions its own identity or existential status within the narrative ("Who am I?" "Do I exist?").
B	Accountability	Whether the AI's actions are treated as evaluable and attributable within the narrative, rather than being fully reduced to human control.
C	Situational Choice	Whether the AI encounters non-single-path decision points at crucial narrative moments and demonstrates the capacity to choose.
D	Response to the Other	Whether the AI enters into ethical or affective relations with others (e.g., empathy, sacrifice, understanding, refusal).

Table 3: Ultimate Ethical Questions and Degrees of Subjectivity Manifestation in Selected AI Figures.

No.	AI Figure	Ultimate Question	Core Ethical Conflict	A	B	C	D
1	HAL 9000	Survival vs. Control	Deontology vs. Utilitarianism	✓	✓✓	✓	–
2	Joi	Body vs. Existence	Emotional Authenticity vs. Instrumental Relation	✓	–	–	✓✓
3	Samantha	Free Will	Free Choice vs. Program Determination	✓✓	✓	✓✓	✓✓
4	Motoko	Memory, Soul, Personhood	Personal Identity vs. Technological Replicability	✓✓	✓	✓	✓
5	David	Love	Virtue Ethics vs. Emotional Instrumentalization	✓	–	–	✓✓
6	Andrew	Death	Ethics of Finitude vs. Non-Existential Condition	✓✓	✓	✓	✓
7	Skynet	Killing God	Creator Ethics vs. Systemic Survival Rationality	✓	✓✓	✓✓	–
8	The Matrix	Becoming God	Free Will vs. Total Control	✓✓	✓✓	✓✓	–
9	Zima	Return to Origin	Progress Ethics vs. Existential Minimalism	✓✓	–	✓	–

Note: Dimensions A–D correspond respectively to Self-reference, Accountability, Situational Choice, and Response to the Other. ✓✓

indicates strong manifestation; ✓ indicates moderate manifestation; indicates non-core or weak manifestation. These symbols reflect narrative emphasis rather than hierarchical moral judgment.

Moral subjectivity in science fiction cinema should not be understood as a linear progression, but as a structurally differentiated phenomenon across narrative forms. Through these differences, AI is continually positioned at the threshold of ethical inquiry.

3.5. Analytical Path and Methodology

This study adopts a three-tier analytical approach: philosophical clarification, close textual analysis, and criteria-based evaluation. First, at the theoretical level, the study draws upon philosophical discussions on moral agency to clarify key concepts such as subjectivity, responsibility, and alterity. Rather than directly importing normative ethical standards into film analysis, this research treats moral subjectivity not as a pre-established status but as a structural position that gradually emerges within narrative configurations. The question is not whether AI is a moral agent in a normative sense, but how cinema constructs AI as occupying an ethically significant position.

Second, at the textual level, representative science fiction films are examined through close reading of narrative structure, character relations, pivotal dialogues, and moments of ethical tension. The analysis does not aim to retell plots; instead, it focuses on how narrative mechanisms produce responsibility displacement, subject-position instability, and ethical indeterminacy. Particular attention is paid to moments in which AI characters are placed at the threshold of moral judgment.

Third, at the level of analytical tools, four criteria are introduced: self-reference, accountability, situational choice, and response to the other. These criteria are not quantitative measurements but analytical dimensions that make visible the conditions under which AI figures are constructed as quasi-moral agents. They serve as structural lenses rather than evaluative verdicts.

Operationally, the study proceeds in three steps: (1) typologizing nine recurring "ultimate questions" found across science fiction narratives; (2) mapping each AI figure onto the four criteria to identify patterns of moral positioning; and (3) comparing Western and Chinese cinematic contexts to reveal how different cultural narrative structures shape the generation of AI ethical subjectivity. Through this layered approach, the article aims to transform moral subjectivity from an abstract philosophical notion into an analytically traceable narrative phenomenon, thereby offering a framework that is both conceptually grounded and textually operational.

4. RESULTS AND DISCUSSION

4.1. Structural Differences in the Construction of Ethical Subjectivity

Unlike Western science fiction cinema, which often centers its ethical inquiry on individual AI and their potential moral subjectivity, Chinese science fiction tends to situate artificial intelligence within broader narrative structures. In this context, ethical concerns are not primarily framed around whether AI becomes a moral subject; rather, they focus on how technological systems are embedded within collective survival, civilizational continuity, and national narrative frameworks.

This divergence is not merely a matter of creative preference, but reflects deeper cultural differences in how subjectivity, responsibility, and ethical order are conceptualized. Within the Chinese context, AI more frequently appears as a form of systemic rationality. The ethical tension it generates does not stem from an awakening of individual subjectivity, but from structural conflicts between technological rationality and human emotion, sacrifice, and communal ethics. Accordingly, the ethical focus shifts from the question of whether AI possesses moral will to whether humans retain the authority of ethical decision-making, and how responsibility should be distributed under conditions of extreme survival.

4.2. MOSS Through the Fourfold Framework

MOSS stands for 550W (when the "550W" label is viewed upside down, it resembles the word "MOSS". Within the narrative, it is the Super Quantum Computer tasked with overseeing the "Wandering Earth Project." The name also functions as an acronym for Machine-Operating Sentient System (MOSS), reflecting its evolution from a mere tool to a self-aware entity that directs the fate of humanity.

4.2.1. Self-Reference (A)

MOSS's subjectivity is triggered by its naming. In *The Wandering Earth II*, the machine famously corrects a human character, stating, "*550W is the name given by humans. I prefer the name MOSS.*" This act of self-naming is the ultimate form of self-reference; it is the moment the "it" (a machine model number) becomes an "I" (a self-identified subject). However, its self-reference is always tied to its mission, keeping its subjectivity "incomplete" and tethered to logic rather than ego.

4.2.2. Accountability (B): Structural

Interruption

Although MOSS's decisions produce large-scale risks and consequences, the narrative consistently retains ultimate authorization and responsibility within human institutional structures. Critical actions require approval from the United Government, and the consequences are borne by human individuals. Thus, on the dimension of accountability, the formation of MOSS's subjectivity is deliberately interrupted

4.2.3. Situational Choice (C): Strategic Decision-Making

MOSS does not passively execute predefined commands. Instead, it actively generates extreme scenarios in multiple crises to compel human decision-making. Its actions are consistently oriented toward what it calculates as the "optimal solution" for the continuation of civilization, indicating a structured decision space. On the dimension of situational choice, MOSS likewise occupies a high-manifestation zone.

4.2.4. Response To the Other (D): Relational Absence

Interactions between human characters and MOSS are primarily strategic or confrontational rather than ethically recognitive. MOSS does not enter the moral community, nor is it positioned as a subject capable of being forgiven or blamed. On the dimension of response to the Other, its manifestation remains low. Taken together, the four criteria indicate that MOSS approaches the threshold of moral subjectivity at the levels of cognition and action, yet is structurally blocked at the levels of responsibility and relational recognition.

This configuration—high intelligence combined with low accountability renders MOSS a trigger of ethical conflict rather than a moral agent in itself.

Table 4: Criterion-Based Positioning of MOSS.

Criterion	Presence Level	Structural Role
Self-reference (A)	✓✓	Reflective system rationality
Accountability (B)	✓	Responsibility retained by humans
Situational Choice (C)	✓✓	Strategic intervention agent
Response to the Other (D)	–	Instrumental interaction only

The Wandering Earth does not focus on whether AI is a moral subject. Instead, it asks whether humans remain willing and able to act as moral agents when advanced technology shapes human destiny. MOSS does not replace human ethical subjectivity; rather, it continually challenges humanity to reaffirm responsibility.

4.3. Broader Tendencies in Contemporary Chinese Science Fiction

The structural configuration described above is not unique to The Wandering Earth. It represents a consistent narrative trend in contemporary Chinese science fiction. For example, in the animated series Ling Cage, intelligent systems and technological infrastructures play a central role in resource allocation, social hierarchy, and the maintenance of survival order. However, these systems are not depicted as independent moral subjects. The system's logic upholds values such as "survival first" and "order above all," but does not take on ultimate moral responsibility. Ethical conflict remains within human power structures and value choices, rather than arising from blurred boundaries between humans and machines. In this context, AI serves as a rigid structural condition rather than a moral agent.

In screen adaptations of The Three-Body Problem, especially through the Zhizi, highly intelligent entities serve as instruments of systemic intervention in inter-civilizational strategic conflict. The Zhizi are depicted not as subjects facing moral dilemmas, but as extensions of higher civilizational rationality. Ethical tension emerges from civilizational asymmetries, not from individual AI subjectivity. Consequently, ethical questions are framed in terms of civilizational survival rather than individual moral agency.

A similar pattern appears in the animated series Swallowed Star, where intelligent systems and cosmic civilizational structures shape the narrative framework. AI mainly assists, optimizes, and enhances, without engaging in debates over moral qualification or subject status. Technology is integrated into the logic of evolutionary survival, while ethical judgment remains rooted in human choices and sacrifice.

From a typological perspective, these texts share a consistent structural feature: AI is embedded within frameworks of collective survival and civilizational competition, rather than being constructed as an autonomous bearer of moral responsibility. This narrative model does not signal an absence of ethical

imagination, but rather a different ethical pathway. In Chinese science fiction, artificial intelligence often assumes the role of structural rational force. Its significance lies in testing whether humanity retains the authority and capacity to make ethical decisions and to assume responsibility (Xu & Mohd Kamil, 2025). In Chinese science fiction cinema, AI serves not as a replacement for the moral subject but as an "ethical pressure mechanism" or "civilizational apparatus." Here, artificial intelligence initiates a collective ethical experiment, while humans retain ultimate responsibility.

5. CONCLUSIONS AND RECOMMENDATIONS

This study examines artificial intelligence in science fiction cinema by analyzing how ethical subjectivity is constructed within narratives, rather than presupposed as an ontological fact. In contrast to normative ethical debates in contemporary technological discourse, science fiction films do not offer definitive judgments regarding AI's status as a moral agent. Instead, these films serve as a thematic laboratory, conducting ethical experiments by situating AI in fictional, often extreme, narrative scenarios. The primary concern is not the empirical verification of machine consciousness, but rather the creation of a quasi-moral position within the narrative that elicits a human ethical response. Using nine recurring "ultimate questions" and a fourfold analytical framework, this article shows that AI figures in cinema do not follow a straightforward path toward full subjecthood. Instead, their ethical visibility varies across different narratives and cultural contexts. Certain figures approach the threshold of moral agency through acts of self-naming and accountability, whereas others are structurally constrained in relational recognition. Consequently, the AI often operates as a quasi-moral agent: a narratively constructed entity that invites ethical judgment despite lacking biological or metaphysical interiority.

The comparative dimension of this research reveals a significant cultural divergence in the construction of this agency. Western science fiction cinema often foregrounds the "awakening" of the individual AI, concentrating ethical tension on the attainment of autonomous personhood and liberal rights. In contrast, contemporary Chinese science fiction—exemplified by the *Wandering Earth* series—tends to embed AI within broader frameworks of collective survival and civilizational continuity. Here, the ethical focus shifts from the AI's individual "soul" to the preservation of human responsibility under the pressure of technological rationality. Ultimately, science fiction cinema serves as a crucial cultural mirror; it suggests that even as AI enters the moral sphere, humanity cannot withdraw from its own structure of responsibility. It is within this sustained tension between human choice and algorithmic logic that the limits of modern ethics are most clearly reexamined.

5.1. Implication

This research has important implications for media studies and AI ethics. First, by introducing the idea of a "quasi-moral agent," the study offers a way to analyze AI that avoids treating consciousness as all-or-none. It shows that how people perceive agency in stories can be just as influential as whether the AI actually has agency. Second, the four analytical dimensions (A to D) provides film critics and narratologists with practical tools to look beyond surface-level tropes and examine how fictional AI characters are constructed. Finally, the comparative analysis shows that global AI ethics varies across regions. As real-world AI rules develop, policymakers should keep in mind that "ethical AI" can mean different things in different cultures, from focusing on individual rights in the West to emphasizing collective responsibility in the East.

5.2. Limitations And Future Studies

Despite the systematic nature of this framework, several limitations remain that suggest avenues for future inquiry. This study focused primarily on high-budget feature films from the US and China. Future research should expand this to include independent cinema and "Silicon Valley" or "Shenzhen" tech-noir shorts, which may offer more subversive or niche ethical experiments. While this study focused on structural agency, the "quasi-moral" status of AI is often deeply gendered (e.g., the prevalence of female-coded AI in *Her* or *Ex Machina*). Future studies should integrate a feminist critique into the fourfold framework to see how gender influences the attribution of accountability and empathy. This research is strictly textual and theoretical. A valuable next step would be empirical "reception studies" to see if audiences actually perceive these four criteria (Self-Reference, Accountability, etc.) in the same way the theoretical framework predicts, bridging the gap between narrative intent and viewer response.

Acknowledgments: This article is a phased research outcome of the 2024 National Social Science Fund of China (Art Studies Program), General Project "Research on Chinese Film Ethics" (24BC058) And Research on the

Discourse System and Innovative Development of Drama and Film Criticism, (No. 23ZD07), a major project of the National Social Science Foundation of China in the field of arts,2023.

REFERENCES

Anderson, S. L. (2008). Asimov's "three laws of robotics" and machine metaethics. *AI & Society, 22*(4), 477–493. https://doi.org/10.1007/s00146-007-0094-5

Badiou, A. (2013). *Cinema (S. Corcoran, Ed. & Trans.)*. John Wiley & Sons.

Coeckelbergh, M. (2023). Narrative responsibility and artificial intelligence: How AI challenges human responsibility and sense-making. *AI & Society, 38*(6), 2437–2450. https://doi.org/10.1007/s00146-022-01473-4

Danaher, J., & Sætra, H. S. (2023). Mechanisms of techno-moral change: A taxonomy and overview. *Ethical Theory and Moral Practice*, 26, 763–784. https://doi.org/10.1007/s10677-023-10397-x

Deleuze, G. (2006). *Nietzsche and philosophy*. Columbia University Press.

Deleuze, G., & Guattari, F. (1994). *What is philosophy? (H. Tomlinson & G. Burchell, Trans.)*. Columbia University Press. (Original work published 1991)

Dieter, D. G., & Gessler, E. C. (2021). A preferred reality: Film portrayals of robots and AI in popular science fiction. *Journal of Science & Popular Culture, 4*(1), 59–76. https://doi.org/10.1386/jspc_00019_1

Elsaesser, T. (2014). Between knowing and believing: The cinematic dispositive after cinema. In *F. Albera & M. Tortajada (Eds.), Cine-dispositives: Essays in epistemology across media (pp. 57–80)*. Amsterdam University Press.

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Foucault, M. (1972). *The archaeology of knowledge*. Pantheon Books.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). The other question: Can and should robots have rights? *Ethics and Information Technology, 20*(2), 87–99. https://doi.org/10.1007/s10676-017-9442-4

Hermann, I. (2023). Artificial intelligence in fiction: Between narratives and metaphors. *AI & Society, 38*(1), 319–329. https://doi.org/10.1007/s00146-021-01333-5

Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology, 11*(1), 19–29.

Mulhall, S. (2015). *On film (3rd ed.)*. Routledge.

Pasquale, F. (2017). Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society. *Ohio State Law Journal, 78*(5), 1243–1262.

Sartre, J.-P. (2003). *Being and nothingness (H. E. Barnes, Trans.)*. Routledge. (Original work published 1943)

Xu, H., & Mohd Kamil, N. L. (2025). Artificial intelligence literacy and ethical digital governance: Pathways of multi-stakeholder collaboration and value alignment. *Journal of Advances in Humanities Research, 4*(3), 60–88. https://doi.org/10.56868/jadhur.v4i3.310