

Rethinking Moral Agency: Luciano Floridi and the Ethics of AI in the Age of ChatGPT

Fransiskus Xaverius Sabu¹ & Taufik Asmiyanto²

^{1,2}University of Indonesia, Indonesia

Correspondence email: fancysabu86@gmail.com

Notes

Submitted: 17-07-2026

Revised: 25-02-2026

Accepted: 12-03-2026

How to cite: Sabu, F. X., & Asmiyanto, T. (2026). Rethinking Moral Agency: Luciano Floridi and the Ethics of AI in the Age of ChatGPT. *Khazanah Al-Hikmah : Jurnal Ilmu Perpustakaan, Informasi, Dan Kearsipan*, 14(1).

<https://doi.org/10.24252/kah.v14i1a5>

DOI: [10.24252/kah.v14i1a5](https://doi.org/10.24252/kah.v14i1a5)

Copyright 2026 © the Author(s)

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



ABSTRACT

Background: As Artificial Intelligence (AI) becomes increasingly pervasive, the technology must be viewed as more than a mere utility for task automation. To preserve human agency, AI integration should facilitate freedom of expression and prevent the systematic marginalization of human roles within the digital ecosystem.

Purpose: This study provides a critical elaboration of Luciano Floridi's Ethics of Information (EOI). The researchers aim to uncover the philosophical essence of Floridi's thought regarding artificial agents and their moral status within the "infosphere."

Methods: Utilizing a comprehensive philosophical literature review and critical analysis, the authors examine Floridi's core tenets. The study maps the transition from traditional anthropocentric ethics to an information-centric moral framework.

Findings: The study reveals that Floridi's reflection on artificial agents necessitates a radical reassessment of moral accountability. By developing the concept of distributed morality, Floridi shifts the ethical burden from a single point of origin (the designer) to a collective, systemic responsibility that encompasses both the technological architecture and its end-users.

Originality/Value: This research highlights a revolutionary ethical model that redefines the relationship between humans and artificial agents. By advocating for a non-anthropocentric approach to moral agency, the study provides a vital framework for ensuring that AI development remains aligned with humanistic values and collective ethical participation.

Keywords: Artificial intelligence; Luciano Floridi; Information ethics

1. INTRODUCTION

Technological advancements have transformed the way humans exist, behave, and interact with the world around them. One of the most influential technological innovations today is artificial intelligence (AI). Since its introduction as a field of research in 1956, AI has had a significant impact across various sectors of human life, including economics, politics, socio-cultural domains, and education. This development has given rise to what is referred to as post-humanism, a condition in which roles previously exclusive to humans can now be increasingly assumed by artificial intelligence (Malčik et al., 2023). AI has driven a new wave of information technology, characterized by information-based systems capable of rapidly

and accurately analyzing data ([Mishra & Pani, 2020](#)). Currently, AI applications can be found in various domains, including chatbots, automated translation tools, robotics, and virtual assistants, which are becoming increasingly integrated into everyday life.

The coexistence of humans and AI has triggered a polarization of interpretations, both in academic discourse and in its real-world applications. The academic debate primarily revolves around ethical considerations: is AI merely a tool that enhances human life, or does it contribute to dehumanization? From an optimistic perspective, AI is seen as a solution that can transcend human cognitive limitations, particularly in solving complex problems that the human mind alone cannot address. One example of its application can be observed in the field of library and information science, where AI is utilized in technical services such as cataloging, library management tasks including staff allocation and decision-making, and user services such as information literacy programs ([Jha, 2023](#)). Meanwhile, in the business sector, AI implementation is expected to drive companies toward greater competitiveness by maximizing their potential and values ([Mishra & Pani, 2020](#)).

The implementation of AI across various aspects of human life brings with it ethical challenges and concerns ([Law et al., 2025](#)). On one hand, the growing need for and dependence on AI continues to rise rapidly, highlighting its increasingly crucial role in daily activities ([Coeckelbergh, 2026](#)). On the other hand, this condition raises a philosophical question regarding the position of humans amid the powerful wave of technological advancement. This anxiety leads to a fundamental dilemma: can AI serve as a tool that empowers humans to reach their fullest potential, or does it instead constrain human creativity and foster greater dependence on technology? ([Rajan, 2024](#)) The technological imperative raises concerns about the risk of dehumanization, wherein humans are reduced to mere technical problems that can be resolved through technical means alone. Therefore, a philosophical reflection on the phenomenon of AI technology is essential to fostering a critical perspective in understanding and evaluating the impact of technology on human existence.

Concerns about the implications of AI technologies for human existence have prompted renewed philosophical reflection on the status of AI as a moral agent. A central question emerging in this debate is whether AI can legitimately be categorized as a moral agent and, if so, whether it should be accorded moral consideration ([Gellers & Holy-Luczaj, 2026](#)). The discussion regarding the moral implications of technological systems can be traced back to Deborah Johnson (2006), who argued that computer systems may be classified as moral entities, though not as moral agents ([Mykhailov, 2021](#)). This distinction rests on the premise that computer systems lack mental states and the volitional capacity necessary for action. At the same time, computer systems may be regarded as moral entities insofar as they are intentionally designed by human creators and operate within contexts structured by human purposes and intentions. In contrast, moral agency presupposes that an actor can be held morally responsible for their actions. Responsibility, in this sense, entails reflective self-control, the capacity to regulate one's conduct and to act on the basis of moral reasons. Such a view aligns with paradigms that conceptualize moral responsibility ([Gogoshin, 2025](#); [Wojewoda, 2025](#)) as grounded in the agent's internal conditions, including self-awareness, intentionality, and rational deliberation – capacities traditionally associated with human beings ([Morandín-Ahuerma, 2025](#)).

Furthermore, debates over the status of artificial agents as moral agents have been sharpened by a range of studies focusing on the question of moral responsibility in relation to artificial intelligence technologies. [Tollon \(2021\)](#) explicitly rejects the claim that moral

agency can only be ascribed to entities possessing consciousness. At a practical level, [Takamatsu \(2025\)](#) highlights the moral dilemmas surrounding the use automated car technology, where human decision makers are often confronted with difficult trade-offs – whether to follow AI-generated recommendation or to adhere the socially accepted moral judgements. In addition, [Telakivi et al. \(2026\)](#) further elaborate the moral status of agency as extended by AI within an interconnected system, arguing that such technological extension can affect the condition under which agents are deemed appropriate bearers of moral responsibility. Their work develops the concept of extended cognition, the idea that the technology may function as a constitutive component of human cognition in ways to analogous to memory and belief. In a similar vein, [Morandín-Ahuerma \(2025\)](#) contends that moral responsibility attributed to non-human systems should be understood as functional responsibility, thereby rejecting essentialist claim that moral agency can only be attributed to conscious subject.

Debates over the moral status of artificial agents can be bridged through Luciano Floridi's notion of distributed morality, which holds that the interconnectivity between humans and non-human entities can give rise to actions that are morally assessable. This arguments rest on an epistemic assumption that agents – whether human or non-human – require information in order to act morally ([Mykhailov, 2021](#)). In this sense, the quality, accuracy, and sufficiency of information significantly shape how agents make morally relevant decisions. Floridi further maintains that an action becomes morally qualifiable insofar as it produces good or harm. As consequence, a technological system such as AI may be regarded as a moral agent when its outputs generate morally significant consequences. Nevertheless, determining whether a system qualifies as a moral agents must be undertaken through the method of abstraction namely Level of Abstraction (LoA). An appropriate LoA may be defined by three criteria: interactivity, autonomy, and adaptability, through which an entity may be classified as a moral agent ([Floridi, 2013](#)). Nevertheless, it provokes a critical question on how to embed structured ethical reasoning into AI, through which human agency could be preserved ([Johnson, 2026](#)).

Discussion on moral agency continue to reveal an epistemic gap regarding how interactions between human and artificial agents may generate actions which can be considered morally. This study seeks to further elaborate the philosophical contributions of Floridi. A crucial conceptual bridge offered by Floridi's information ethics provides a critical framework for examining artificial intelligence and its implications for human existence. Thus, this study explores Floridi's theoretical position in response to the increasingly pervasive presence of AI, particularly within educational context marked by the widespread adoption of systems such as those developed by OpenAI. A deeper philosophical engagement with Floridi's concept is expected to contribute to a moral balanced and theoretically grounded discourse on the ethical and civilizational implications of contemporary AI technologies.

2. METHODS

This study employs a qualitative philosophical approach supported by a structural literature review. The research is designed to critically examine the relevance of Luciano Floridi's information ethics to contemporary developments in artificial intelligence, particularly generative AI systems such as ChatGPT. Rather than treating Floridi's work descriptively, the study operationalizes key concepts through a systematic analytical procedure that combines conceptual and interpretative synthesis.

The primary corpus consists of two of Floridi's seminal works – *The Philosophy of Information* (2011) and *The Ethics of Information* (2013) – which are treated as the core theoretical sources. These works were selected because they articulate Floridi's most comprehensive formulation of information ethics, including his account of moral agency, distributed morality, and the Level of Abstraction (LoA). To broaden the analytical perspective and situate Floridi's position within ongoing scholarly debates, the study also incorporates peer-reviewed journal articles that explicitly engage with Floridi's framework of information ethics and artificial moral agency.

The selection of secondary literature followed explicit inclusion criteria. Sources were included if they (1) directly discuss Floridi's information ethics or moral agency, (2) demonstrate conceptual relevance to AI ethics or artificial agents, and (3) are published in peer-reviewed academic venues. Works that discuss AI ethics without substantive engagement with Floridi, non-academic publications, and purely technical AI papers were excluded. This filtering process was intended to ensure conceptual coherence and analytical depth.

Data analyses proceeded through four interrelated stages. First, conceptual extraction was conducted through iterative close reading of the primary texts. At this stage, key construct – such as distributed morality, moral agency, and the Level of Abstraction – were systematically identified and coded. The purpose of this phase was to generate a structure inventory of Floridi's core conceptual apparatus. Second, the extracted concepts were organized through thematic categorization. Using an axial coding strategy, related concepts were grouped into higher order themes, including: the ontological status of artificial agents, conditions for moral agency, and the informational foundation for moral action. This step enabled the reconstruction of the internal architecture of Floridi's ethical framework. Third, the study employed comparative interpretation of critically situate Floridi's position within the broader discourse on AI ethics. At this stage, interpretation derived from the primary texts were juxtaposed with arguments found in the selected secondary literature. Finally, the reconstructed conceptual framework was analytically applied to contemporary case of generative AI systems, particularly to ChatGPT-like large language models. The analysis examined whether such systems satisfy Floridi's criteria for artificial agency – interactivity, autonomy, and adaptability – when evaluated at an appropriate Level of Abstraction. Throughout the study, analytical rigor was maintained through transparent coding procedure, iterative reading, and careful alignment between conceptual categories and textual evidence.

3. RESULTS AND DISCUSSION

Luciano Floridi and His Contribution to the Philosophy of Technology

Luciano Floridi is an Italian-born philosopher (b. 1964) renowned for his contributions to the fields of information ontology, information epistemology, information ethics, and information logic. He currently serves as the Director of the Digital Ethics Center at Yale University and professor of Communication and Cultural Sociology at the University of Bologna. In 2020, Scopus recognized Floridi as the most frequently cited contemporary philosopher in academic publications.

The philosophical discourse developed by Floridi on technology revolves around three main topics: the impact of technological development on social, political, and cultural

aspects; the epistemological status of knowledge related to technology; and the ontological status of technological artifacts (Demir, 2012). Floridi's conceptual framework for analyzing technology is rooted in his ideas on information epistemology. According to Floridi, information is data that has form, meaning, and truth. Information is not merely understood as a representation of reality, but rather, it is reality itself. Through rationality, humans comprehend information not only for the sake of their intellect but also as a self-defense mechanism for survival (Floridi, 2011). Therefore, understanding the epistemic status of technological development is crucial, and in Floridi's thought, it has shifted from a materialist approach to an informational approach.

Floridi is engaged in an academic discourse on the existence of technology and its impact on human life by emphasizing that the development of information and communication technology has altered the way we perceive the human self-concept and how individuals interact with others (Floridi, 2011). This analysis begins with a fundamental question, "What's next?" to highlight his philosophical stance as a bridge between optimism and skepticism regarding technology. The essence of technology within the sphere of human civilization is understood as a mediator between humans and their world. According to Floridi, the function of technology as a medium connecting to humans with the physical world has undergone three crucial phases. First, technological artifacts served as a direct intermediary between humans and nature; second, technology became a provider of energy for other technologies; and third, technology evolved into a system that connects various technologies. The transformation of technology in the third phase, as seen in the Internet of Things (IoT), has led to a significant leap in human civilization's history. However, this shift simultaneously gives rise to new challenges, namely the increasing marginalization of the human role within the technological constellation (Rajan, 2024).

Within the framework of the philosophy of information, Floridi argues that the human being is an *informational organism* (*infor*g), a conscious and interconnected agent situated within a broader *infosphere*. The *infosphere*, a neologism coined by Floridi, denotes a complex informational ecosystem encompassing data, communication, and the dynamic interplay between humans and technology, in which the boundaries between the online and offline worlds become increasingly blurred. This hybrid realm is what Floridi terms the *onlife* dimension. As an informational ecosystem, the *infosphere* not only offers a space of connectivity for humans as informational subjects, but also enables the emergence of artificial agents, non-human entities endowed with a degree of autonomy, self-regulation, and purposiveness. The conceptual architecture of the informational space proposed by Floridi is profoundly influenced by Kant's transcendental categories, particularly the notion of space. Kant reconfigured the classical debate between Newton's conception of absolute and real space and Leibniz's view of ideal and relative space by introducing the idea of space as both ideal and absolute. Building upon this, Floridi constructs the idea of the informational space as ideal in a Kantian transcendental sense and as relative in the Leibnizian sense. It is ideal insofar as it serves as a condition of possibility for the experience of a particular system; it is relative insofar as it does not exist independently from the informational objects that an epistemic agent may encounter. Thus, the ideality and relativity of the informational space further articulate its epistemological nature as an ordered space, grounded in the relationality among informational objects (Durante, 2017).

The philosophy of information constructed by Floridi not only addresses the concepts of information and computation but also elaborates on its methods and models in relation to philosophical issues. One of the key concepts in Floridi's philosophy is the concept of the re-

ontologization of information space in technology, specifically the transition from analog to digital data, which enables data to be stored, accessed, and manipulated quickly and efficiently (Floridi, 2013). This transition has led to the widespread growth of information space, marked by the escalation of the convergence of digital resources and devices. The ontology or essence of information technology (such as software and algorithms) becomes increasingly compatible with its object, which is the data being processed. This means that what is referred to as the digital process is not a singular, standalone process but rather a unity between the information processor (the computer) and the information being processed (the data). Consequently, information space becomes increasingly free from ontological friction, thereby facilitating the flow and access to information. Floridi emphasizes that this condition brings with it several consequences, including the substantial erosion of the right to ignore, exponential growth in shared knowledge, and the increasing responsibility of agents (Floridi, 2013).

The Phenomenon of Artificial Intelligence

Artificial Intelligence (AI) is conceived as a computational system endowed with cognitive faculties, designed to emulate human capacities such as problem-solving, decision-making, and experiential learning (Dursun & Bilici Geçer, 2024). At its core, this definition gestures toward three interrelated dimensions, system, cognition, and behavior, that collectively render AI not merely a technological artifact, but a functional agent capable of engaging with the complexities of human life. The notion of intelligence attributed to such systems suggests more than mere algorithmic efficiency; it invokes a form of embedded rationality, wherein AI appears to exhibit a degree of autonomous reasoning and intentional action (Calzati, 2023; Wojewoda, 2025). This raises fundamental philosophical questions about the nature of mind, agency, and the evolving boundaries between the human and the artificial.

As a manifestation of advances in information technology, Artificial Intelligence (AI) encompasses several conceptual forms (Saveliev & Zhurenkov, 2021):

Table 1. Types of Artificial Intelligence Systems

No.	Type of AI	Description
1	Narrow AI	Refers to software designed to perform specific tasks, such as complex computations, object recognition, automated language translation, and gameplay.
2	Autonomous AI	Describes systems capable of operating independently for extended periods without direct human intervention.
3	Adaptive AI	Denotes systems that can adjust to new conditions and autonomously enhance or develop their own capabilities over time.
4	Artificial General Intelligence (AGI)	Represents highly flexible AI systems that can perform a wide range of tasks, either independently or in collaboration with human operators.
5	Strong AI	Refers to systems whose cognitive and adaptive capacities are comparable to those of the human mind.
6	Super-Human AI	Describes AI systems whose operational and intellectual capabilities surpass human intelligence.

The most commonly developed AI models today are *Machine Learning* (ML) and *Large Language Models* (LLMs), which primarily rely on *Natural Language Processing* (NLP). These technologies enable AI systems to comprehend, interpret, and generate text in ways that closely resemble human linguistic capabilities. In recent years, public discourse surrounding AI has intensified, particularly with the emergence of *generative AI*, as exemplified by applications such as ChatGPT developed by OpenAI, Microsoft's Bing, Google's Bard, and AI-integrated features across Meta's suite of social media platforms.

Artificial Intelligence (AI) has exerted a profound influence across multiple sectors, including industry, education, healthcare, and communication. In the industrial domain, AI plays a pivotal role in technological transformation, particularly in enhancing decision-making processes through decision support systems. In this context, AI is employed to optimize supply chains, improve the efficiency of manufacturing procedures, and refine various operational dynamics (Soori et al., 2024). The integration of AI into industrial systems reveals a striking capacity for large-scale data processing, precise pattern recognition, and the automation of increasingly sophisticated and adaptive processes. Beyond its technical capabilities, the presence of AI in these domains invites reflection on the shifting locus of agency and judgment, where decisions once rooted in human experience are now increasingly mediated by algorithmic reasoning.

The rapid development of Artificial Intelligence (AI) is largely driven by advances in algorithmic design and the exponential growth of big data. The widespread use of digital technologies has enabled technology industries to collect vast quantities of data, which now serve as the primary resource for training and evolving AI systems (Calzati, 2023). As more individuals engage with digital devices and applications, the volume of personal data harvested by platform providers increases correspondingly. This dynamic gives rise to the phenomenon of *datafication*, a condition in which human activities are rendered into quantifiable, digital traces that can be accessed, analyzed, and processed by computational systems. At a deeper level, datafication marks a shift in how human existence is conceptualized: from embodied, lived experience to a series of data points, thus inviting critical reflection on the ontological and ethical implications of reducing human life to algorithmically legible formats.

Floridi's Thought in Responding to The Ethical Challenges of AI

Before delving into Luciano Floridi's ethical reflections on Artificial Intelligence (AI), it is essential to first grasp the foundational concept of AI ethics. According to Leslie, AI ethics refers to a set of values, principles, and techniques grounded in widely accepted standards that serve as normative guidelines for the development, deployment, and moral decision-making related to AI technologies (Christoforaki & Beyan, 2022). This definition underscores that AI ethics is not merely concerned with the moral implications of AI usage, but also with embedding ethical considerations from the earliest stages of design and creation through to real-world application. As such, the study of AI ethics is intimately connected to the philosophy of technology, particularly in examining the ethical dimensions of computational systems and machines, core components in the construction of AI. This intersection invites a broader philosophical inquiry into how technological artifacts shape, and are shaped by, human values and moral responsibility.

Floridi's thought, as articulated in his philosophy of *information ethics*, can be understood as a call for moral responsibility in the face of technological advancement and transformation. In his seminal work *The Ethics of Information*, Floridi asserts that moral action

within the framework of information ethics is best conceived as a pattern of information processing, wherein information itself is an abstraction of the broader ecosystem (Floridi, 2013). In this view, the adequacy and integrity of information become crucial elements in forming ethical evaluations of agents. However, Floridi's notion of information ethics should not be mistaken for a singular ethical theory; rather, it constitutes a broader research program, a form of *macroethics*, that concentrates on the analysis of informational processes within the larger communicative chain (Demir, 2012). This perspective invites a reconfiguration of traditional moral philosophy by situating ethical agency within the relational and systemic dynamics of the *infosphere*.

A fundamental question in the ethical analysis of Artificial Intelligence (AI) concerns the very notion of moral agency: what qualifies as a moral agent, and how might artificial agents be situated within a moral framework? According to Floridi, a moral agent is a system that is interactive, autonomous, and adaptive, capable of performing actions that are subject to moral evaluation (Floridi, 2013). This definition implies that, in principle, artificial agents do not possess the inherent capacity to act morally in the human sense. However, Floridi allows for the possibility that such capacity might be eventually developed or approximated in artificial systems. Consequently, he proposes that while artificial subjects lack the cognitive intelligence necessary for moral responsibility, they may nevertheless be regarded as sources of moral action. In this way, Floridi reframes the ethical discourse not in terms of inner intentionality or consciousness, but in terms of the systemic and relational effects of agency within the broader *infosphere*.

To understand how ethical principles operate within the informational domain, it is essential to examine Luciano Floridi's concept of the *Level of Abstraction* (LoA). This concept functions as an ontological perspective that facilitates the processing of information and interaction with the world, thereby enabling a wide range of philosophical analyses, particularly within the domains of information ethics and the philosophy of information (Floridi, 2013). The notion of LoA is derived from modeling practices in computer science, specifically from what is known as the *Formal Method*, a mathematical modeling technique used to define and verify the behavior of computational systems, ensuring their systematic design and functionality (Floridi, 2011). Within this framework, variables in a model correspond to observable aspects of reality. Floridi's formulation of abstraction is deeply informed by Immanuel Kant's transcendental philosophy, which holds that human knowledge is limited to phenomena that are accessible to the senses and interpreted through the mind's inherent cognitive structures, or categories. As such, any abstraction method necessarily yields analyses that are shaped by the principles of causality and modality, which are themselves contingent upon how a system is understood at a given level of abstraction (Floridi, 2013). In this light, the LoA becomes not merely a technical tool, but a philosophical lens through which the ethical dimensions of information systems may be rigorously interrogated.

The concept of *Level of Abstraction* (LoA) carries significant philosophical implications across a range of domains, including agency, the Turing test, perception, quantum observation, artificial life, and simulation. In this discussion, the focus is placed on a reflective analysis of its implications for the concepts of agency, artificial life, and simulation. An entity may be identified as an *agent* if it possesses the capacity to respond to external stimuli in a manner that is interactive, autonomous, and adaptive (Floridi, 2011). From a technological standpoint, a software system may be classified as an agent insofar as it demonstrates these operational characteristics, interaction with its environment, autonomous functioning, and

adaptability to changing conditions. On the basis of these assumptions, Artificial Intelligence (AI) is accordingly categorized as an *artificial agent*, given its capacity to operate within these parameters. This classification prompts deeper philosophical inquiry into the nature of agency itself, as well as the ontological status of synthetic entities that increasingly blur the boundaries between the organic and the artificial, the simulated and the real.

Within the categorization of artificial agents, AI *in se*, according to Floridi, does not possess the capacity to act morally. The moral status of an agent is traditionally evaluated based on its ability to exercise responsibility and freedom. Floridi contends that an action qualifies as morally relevant if it results in moral good or evil, that is, if it contributes to an increase or decrease in the metaphysical entropy of the *infosphere* (Floridi, 2013). This conceptualization reframes moral value not in terms of internal intentions, but in terms of systemic consequences within the informational environment. Within this framework, AI cannot be considered a moral agent for several reasons: it lacks intrinsic purpose, it possesses no consciousness, it is devoid of freedom, and it bears no responsibility for the outcomes of its operations. In other words, while AI may be the source of morally significant actions, it cannot itself be the bearer of moral agency. This distinction underscores a critical boundary between artificial and human agents, one grounded not merely in function, but in ontological and ethical depth.

Within Floridi's philosophy of information, often interpreted as a form of relational ontology, the notion of moral action is conceived as the outcome of complex interactions among distributed systems, a framework he terms *distributed morality*. This perspective implies that moral actions are not solely attributable to individual agents, but rather emerge from the collective agency of multiple entities, both human and artificial, operating within a broader and more intricate socio-technical network. Such a view carries significant implications for debates concerning moral responsibility. From Floridi's standpoint, technological responsibility must be understood as inherently distributed. Stahl et al., (2023) reinforce this position by examining artificial agents endowed with the capacity for autonomous learning and behavior. They argue that the presence of such capacities does not negate the enduring moral accountability of those who design and implement these systems. On the contrary, recognizing artificial agents as moral agent risks diminishing, or even displacing, human responsibility for the design, deployment, and consequences of these technologies. Thus, Floridi's approach warns against the ethical hazard of outsourcing responsibility to the very systems we create, emphasizing instead the persistent need for human oversight and ethical deliberation in the age of autonomous technologies.

Ontology, in Floridi's framework, is not to be understood as a metaphysics of *Being* in the traditional sense, but rather as a conceptual tool for apprehending reality, one in which an agent is capable of mastering, attributing meaning to, and conceptualizing its environment according to specific *levels of abstraction* (Floridi, 2013). This claim carries significant implications for the philosophical problem of communication within informational environments. Floridi argues that communication becomes possible when interacting entities operate at the same level of abstraction, thereby enabling them to share an ontology, what he terms *shared ontology*. However, this proposition provokes a deeper philosophical question: does the mere presence of a shared ontology between two entities suffice to ensure the possibility of genuine communication? In other words, is ontological alignment a sufficient condition for communicative understanding, or are there other cognitive, interpretive, or semantic dimensions that must also be accounted for? Such inquiries open up

critical terrain for rethinking the nature of meaning, understanding, and dialogue in both human and artificial contexts.

The key concepts offered by Floridi as analytical tools for examining the ethical questions surrounding the existence of AI technologies are not confined to the realm of theoretical discourse. In what follows, the authors seek to articulate Floridi's philosophical insights in an applied manner, as a normative guide for the ethical formulation of policies concerning the development of artificial intelligence technologies.

The Application of Floridi's Perspective of AI Development Policy

Technological advancement does not merely bring about positive developments in human life; it also gives rise to a range of profound challenges that significantly shape the trajectory of human civilization. One of the most pressing ethical concerns associated with the use of artificial intelligence is its disruptive potential, namely, its tendency to undermine human control over technological systems (Khan et al., 2022). For this reason, ethical discourse on AI must not be confined to the academic sphere alone but must also extend into the domain of practical application. At the academic level, debates surrounding AI ethics revolve around the construction of a robust philosophical foundation capable of comprehensively addressing the ontological status of AI within the broader context of human reality. In the practical domain, AI ethics serves as a normative guide for both the design and deployment of artificial intelligence technologies, ensuring that their development remains aligned with human values and responsibilities.

Ethical evaluation in the creation and deployment of artificial intelligence draws upon several philosophical frameworks, including utilitarianism, deontological ethics, virtue ethics, and discourse ethics (Christoforaki & Beyan, 2022). From the standpoint of utilitarianism, whose principal figures include Jeremy Bentham and John Stuart Mill, moral judgment is grounded in the principle of utility, wherein the ethical value of an action is determined by the consequences it produces, particularly in terms of maximizing overall well-being or minimizing suffering. This consequentialist framework invites a cost-benefit analysis of AI technologies, urging developers and policymakers to assess not only the efficiency and effectiveness of AI systems, but also their broader impact on human flourishing.

The dialectics of AI implementation in human life bears significant implications for social inequality, potentially exacerbating disparities experienced by marginalized groups within society. AI systems may operate on biased or flawed algorithms, resulting in decisions that lack transparency and defy accountable justification (Kieslich et al., 2022). Accountability thus emerges as a crucial ethical category through which human attitudes toward the potentially destructive consequences of AI systems must be examined. In this context, the ethical framework of information, as articulated by Luciano Floridi, becomes profoundly relevant. His thought offers a philosophical lens through which the deployment of AI within the intricate fabric of human reality, particularly in the domain of public policy, may be ethically evaluated and normatively guided.

First, Floridi's concept of informational privacy may be interpreted as an ethical guideline for designing human-centered AI policy. This concept is articulated through what he terms the function of *ontological friction* within the infosphere (Floridi, 2013). Floridi argues that the advent of AI technologies has significantly contributed to the reduction of such friction, namely, the resistance between informational entities, by enabling artificial agents to access, process, and disseminate information with minimal human oversight and regulation. Informational friction refers to the barriers that affect the flow of information

within an environment, which in turn influence the degree of privacy afforded to agents based on the accessibility of their personal data (Floridi, 2014). Furthermore, informational friction plays a constructive role in slowing the diffusion of information, mitigating risks of misuse, and reinforcing mechanisms for verifying the authenticity and integrity of data.

Furthermore, the understanding of informational privacy as a function of ontological friction is closely related to the concept of accessibility. Within the infosphere, there operates a postulate: the greater the degree of informational accessibility, the more difficult it becomes to monitor and safeguard informational privacy, particularly with regard to personal data. This notion offers a critical point of reference for the formulation of policies aimed at protecting privacy. As previously discussed, AI technologies function optimally when supported by vast and robust data infrastructures; therefore, policies concerning the use of personal data must be carefully designed and guided by ethical considerations that prioritize human well-being. Ethically grounded policies on personal data usage must be adopted by AI developers, with the goal of fostering technological systems that uphold and promote human dignity.

In alignment with the problem of informational privacy, the development of AI technologies grounded in big data is intrinsically linked to issues of transparency (Walker et al., 2025), particularly in the implementation of algorithms that are free from bias and discrimination (Kieslich et al., 2022). The research conducted by Kieslich et al. (2022) clearly identifies the discriminatory potential embedded within certain AI features, often directed at specific social groups, a consequence of flawed and biased data input processes. In this context, Floridi's proposal for ethical responsibility offers a valuable moral compass to guide the formulation of AI development policies. Floridi decisively rejects the traditional view that assigns primary responsibility to the software developers or coders for the ethical consequences of AI systems. Instead, he proposes that accountability in AI technologies emerges from a distributed collaboration among multiple actors, including interface designers, access protocol engineers, and data traffic architects (Floridi, 2013). Such a distributed model of responsibility reflects a more nuanced and systemic approach to ethics in the age of artificial agency. To strengthen this argument, a model a responsibility that emphasize shared responsibility must be develop consistently, particularly by embedding ethical principal into artificial intelligence systems. Accordingly, there is a pressing need for an ethical framework that can be translated into algorithmic form so that it can be processed by artificial intelligence technologies (Dodig-Crnkovic et al., 2025).

Secondly, we encounter the paradox of artificial intelligence: agency without intelligence. In 2023, Floridi authored a critical reflection bearing the provocatively titled *AI as Agency without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models*. In this work, Floridi argues that ChatGPT, an exemplary application of AI built upon the architecture of Large Language Models (LLMs), constitutes a fragile system which, at various points, generates information that is invalid and epistemically unaccountable (Floridi, 2023). Within its mechanistically designed operational framework, the LLM serves primarily as a massive data aggregator that processes linguistic input according to user prompts. While such models demonstrate the capacity to generate semantically coherent content, they nevertheless exhibit significant limitations: a lack of logical inferential capacity, insufficient contextual learning, and a pronounced susceptibility to error. When faced with queries beyond its trained corpus or interpretive capacity, ChatGPT is prone to generating fictitious responses, often embedding inaccurate information and fabricated references. This

epistemological fragility reveals the core tension in LLM-based systems, namely, that they simulate agency without possessing genuine intelligence or understanding.

In addressing the widening gap between the implementation of AI technologies and the ethical imperatives that necessarily accompany them, Floridi advocates the adoption of the *Association for Computing Machinery* (ACM) Code of Ethics as a normative framework to ensure a balance between technological innovation and human-centered ethical standards. This code offers a set of guiding principles for AI designers and developers, aiming to foster the creation of systems that serve the flourishing of human beings rather than reducing them to mere objects of computation. According to (Floridi, 2013), the following ethical commitments are essential:

1. To contribute to society and the well-being of humanity
2. To avoid harm to others
3. To be honest and trustworthy
4. To act fairly and without discrimination
5. To respect property rights, including copyrights and patents
6. To give proper credit for intellectual property
7. To respect the privacy of others
8. To honor confidentiality

These principles serve not merely as regulatory mechanisms but as ethical anchors, reaffirming the primacy of human dignity in the age of increasingly autonomous artificial agents.

4. CONCLUSION

Luciano Floridi's conception of information ethics, when applied to the dilemmas posed by artificial intelligence (AI), may be understood as a form of meta-ethics, one that seeks to establish a philosophical foundation for the development and implementation of AI technologies oriented toward the flourishing of humanity. By positioning AI as an artificial agent within a moral evaluative framework, Floridi transcends the boundaries of classical ethical paradigms, which tend to restrict moral consideration to human entities alone. Within the *infosphere*, AI is not merely an aggregator of information but ought also to function as a conduit through which information flows responsibly, upholding the values of freedom, accountability, and human dignity. From this reflection, we may conclude that ethical responsibility in the age of AI is fundamentally distributed. It does not rest solely upon system designers and developers but also extends to the users who engage with AI technologies in everyday life. As morally conscious agents, AI users bear an ethical obligation to engage with such technologies freely, wisely, and with steadfast commitment to the preservation of human dignity.

AUTHORS' CONTRIBUTIONS

Fransiskus Xaverius Sabu: Writing original draft preparation. Ideas. Methodology.
Taufik Asmiyanto: Supervision.

CONFLICT OF INTERESTS

We state that there are no known conflicts of interest linked with this publication, and that there has been no significant financial assistance for this work that could have influenced its outcome.

REFERENCES

- Calzati, S. (2023). From big data epistemology to AI politics: rescuing the public dimension over data-driven technologies. *Journal of Information, Communication and Ethics in Society*, 21(3), 358–372. <https://doi.org/10.1108/JICES-12-2022-0108>
- Christoforaki, M., & Beyan, O. (2022). AI Ethics, A Bird's Eye View. *Applied Sciences (Switzerland)*, 12(9). <https://doi.org/10.3390/app12094130>
- Coeckelbergh, M. (2026). AI and Epistemic Agency: How AI Influences Belief Revision and Its Normative Implications. *Social Epistemology*, 40(1), 59–71. <https://doi.org/10.1080/02691728.2025.2466164>
- Demir, H. (Ed.). (2012). *Luciano Floridi's Philosophy of Technology: Critical Reflections* (H. Demir, Ed.). Springer. <http://www.springer.com/series/8657>
- Dodig-Crnkovic, G., Basti, G., & Holstein, T. (2025). Delegating Responsibilities to Intelligent Autonomous Systems: Challenges and Benefits. *Journal of Bioethical Inquiry*, 22(3), 507–514. <https://doi.org/10.1007/s11673-025-10428-5>
- Durante, M. (2017). Ethics, Law and the Politics of Information: A Guide to the Philosophy of Luciano Floridi. In *International Library of Ethics, Law and Technology* (Vol. 18).
- Dursun, D., & Bilici Geçer, R. (2024). Can artificial intelligence models serve as patient information consultants in orthodontics? *BMC Medical Informatics and Decision Making*, 24(1). <https://doi.org/10.1186/s12911-024-02619-8>
- Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.
- Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.
- Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models. In *Philosophy and Technology* (Vol. 36, Number 1). Springer Science and Business Media B.V. <https://doi.org/10.1007/s13347-023-00621-y>
- Gellers, J. C., & Holy-Luczaj, M. (2026). Consider the xenobot: moral status for intelligent machines revisited. *Law, Innovation and Technology*. <https://doi.org/10.1080/17579961.2025.2593778>
- Gogoshin, D. L. (2025). A way forward for responsibility in the age of AI. *Inquiry (United Kingdom)*, 68(4), 1164–1197. <https://doi.org/10.1080/0020174X.2024.2312455>
- Jha, S. K. (2023). Application of artificial intelligence in libraries and information centers services: prospects and challenges. *Library Hi Tech News*, 40(7), 1–5. <https://doi.org/10.1108/LHTN-06-2023-0102>
- Johnson, J. (2026). Can AI behave ethically during military crises? Preserving human moral agency. *International Affairs*, 102(1), 63–83. <https://doi.org/10.1093/ia/iaf191>
- Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A Systematic Literature Review of Principles and Challenges. *ACM International Conference Proceeding Series*, 383–392. <https://doi.org/10.1145/3530019.3531329>
- Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data and Society*, 9(1). <https://doi.org/10.1177/20539517221092956>
- Law, R., Ye, H., & Lei, S. S. I. (2025). Ethical artificial intelligence (AI): principles and practices. *International Journal of Contemporary Hospitality Management*, 37(1), 279–295. <https://doi.org/10.1108/IJCHM-04-2024-0482>
- Malčík, M., Rakowski, R., Miklošiková, M., Zemčík, T., & Feber, J. (2023). Philosophical Background of Computational Thinking. *International Journal of Emerging Technologies in Learning*, 18(17), 126–135. <https://doi.org/10.3991/ijet.v18i17.41139>

- Mishra, A. N., & Pani, A. K. (2020). Business value appropriation roadmap for artificial intelligence. In *VINE Journal of Information and Knowledge Management Systems* (Vol. 51, Number 3, pp. 353–368). Emerald Group Holdings Ltd. <https://doi.org/10.1108/VJKMS-07-2019-0107>
- Morandín-Ahuerma, F. (2025). Freedom Without Conscience: Computational Compatibilism and Moral Attribution in Artificial Intelligence Systems. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2025.2565393>
- Mykhailov, D. (2021). A moral analysis of intelligent decision-support systems in diagnostics through the lens of Luciano Floridi's information ethics. *Human Affairs*, 31(2), 149–164. <https://doi.org/10.1515/humaff-2021-0013>
- Rajan. (2024). Philosophy of technology for the lost age of freedom: a critical treatise on human essence and uncertain future. *AI and Society*. <https://doi.org/10.1007/s00146-024-01942-y>
- Saveliev, A., & Zhurenkov, D. (2021). Artificial intelligence and social responsibility: the case of the artificial intelligence strategies in the United States, Russia, and China. *Kybernetes*, 50(3), 656–675. <https://doi.org/10.1108/K-01-2020-0060>
- Soori, M., Jough, F. K. G., Dastres, R., & Arezoo, B. (2024). AI-Based Decision Support Systems in Industry 4.0, A Review. *Journal of Economy and Technology*. <https://doi.org/10.1016/j.ject.2024.08.005>
- Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. <https://library.oapen.org/handle/20.500.12657/59315%0Ahttps://library.oapen.org/bitstream/20.500.12657/59315/1/978-3-031-17040-9.pdf>
- Takamatsu, R. (2025). With a Little Help from AI: Moral Orientation, Trust in AI, and Acceptance of AI-Based Advice Utilization in Moral Dilemmas Involving Partially Automated Car. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2025.2531285>
- Telakivi, P., Kokkonen, T., Hakli, R., & Mäkelä, P. (2026). AI-Extended Moral Agency? *Social Epistemology*, 40(1), 116–128. <https://doi.org/10.1080/02691728.2025.2472759>
- Tollon, F. (2021). Do others mind? Moral agents without mental states. *South African Journal of Philosophy*, 40(2), 182–194. <https://doi.org/10.1080/02580136.2021.1925841>
- Walker, P. B., Haase, J. J., Mehalick, M. L., Steele, C. T., Russell, D. W., & Davidson, I. N. (2025). Harnessing Metacognition for Safe and Responsible AI. *Technologies*, 13(3). <https://doi.org/10.3390/technologies13030107>
- Wojewoda, M. (2025). The Ethics of Responsibility in the Context of the Use of Intelligent Machines and the Problem of the Technosystem. *Forum Philosophicum*, 30(2), 199–213. <https://doi.org/10.35765/forphil.2025.3002.10>