MELINTAS
40.3.2024 [336-358]

THE CENTRAL ROLE OF *IN-FORMATION* IN THE AGE OF ARTIFICIAL INTELLIGENCE AND ITS SIGNIFICANCE TO LEGAL PERSONHOOD

Mardohar B.B. Simanjuntak[1] | Reader
Bandung, Indonesia

Tristam P. Moeliono[2] | Faculty of Law
Parahyangan Catholic University
Bandung, Indonesia

Abstract:
Artificial intelligence (AI) can be understood from an epistemological standpoint, specifically in relation to the concept of legal personhood. It has evolved into a stronger version with broader capabilities, surpassing its predecessors, particularly Artificial Narrow Intelligence (ANI). This rapid advancement has significant consequences, not to mention its increasing problem-solving ability–an essential characteristic of intelligence. To determine its legal status, AI must meet specific criteria. Generative AI, such as large language models (LLMs), serves as a warning sign, as Yuval Noah Harari argues, radically altering how humans cooperate. Harari maintains that social constructs have historically relied on human-to-document intersubjective interaction, mediated by bureaucracies. The latest developments, however, shift this relationship toward document-to-document interaction, transforming the nature of information. Instead of being merely a collection of data ('information'), information now serves as the fundamental backbone of social institutions ('in-formation'). From a periscopic perspective, as discussed

by the authors, the discrepancy between the real and the metaversal expands AI's role in everyday human activities. The authors propose a special legal personhood status for AI, considering its emerging analytic a priori capabilities. However, while AI remains in the later stages of its development, this possibility introduces mitigated risks that require further analyses.

Keywords:
artificial intelligence • *legal personhood* • *intersubjective interaction* • *periscopic view* • *mitigated risks*

Introduction

The concepts of will theory and interest theory serve as foundational pillars in understanding the epistemological ground of legal personhood. Will theory posits that legal rights arise from an individual's ability to exercise will, autonomy, and agency. This approach emphasizes the importance of personal autonomy in shaping legal obligations and entitlements. In contrast, interest theory suggests that legal rights exist to protect individual or collective interests, even when the capacity to exercise will is absent. These two theories intersect in profound ways, particularly when applied to questions of legal personhood and the broader epistemological framework underpinning our understanding of knowledge and decision-making (Kurki, 2019:32-39).

Legal personhood depends intricately on the interplay between will theory and interest theory. Traditionally, legal systems have conferred personhood based on an entity's capacity for will and rationality, as exemplified by human beings. However, the inclusion of entities such as corporations as in *personae ficta* challenges this notion, as their personhood is rooted in interest theory rather than individual autonomy. Corporations act as entities with collective interests rather than autonomous wills. This intersection raises critical epistemological questions: how do we determine who or what qualifies as a legal person, and how does our understanding of knowledge influence this determination?

The epistemological underpinnings of will and interest theories can be explored through a tripartite structure. The first part of this inquiry addresses the nature of knowledge itself, asking, "How do we know what we know?" Sentience, as explored by Yuval Noah Harari in *Nexus*, emerges as a critical element in this discussion. Sentience underpins our ability to perceive, reason, and decide, framing human beings as unique epistemological agents. Harari's concept of sentience connects to the role of *personae ficta* rooted in informational worlds, which embody constructed legal identities that lack sentience but are still recognizable within progressive legal frameworks.

The second part delves into the parameters of knowledge, distinguishing between analytic and synthetic knowing. Analytic epistemic approach involves deductive reasoning and logical structures, while the synthetic one incorporates empirical observation and experience. Harari's theory highlights how the use of information defines both human and artificial systems (Harari, 2024:195-200). For humans, virtues and knowledge are encoded in scriptures and cultural traditions, regarded as infallible foundations of communication and decision-making. However, as artificial intelligence (AI) becomes increasingly sophisticated, it redefines the infallibility of information. AI's strength lies in its ability to muster vast informational resources and process them with unparalleled speed and precision. This capability challenges human limitations in accessing and utilizing knowledge.

The infallibility of information, once considered the bedrock of human decision-making, is being recontextualized in the AI era. AI systems, designed to operate on massive data sets and probabilistic models, offer a form of infallibility distinct from traditional human sources. However, the decentralization of this informational power raises ethical and philosophical dilemmas. While humans rely on sentience and consciousness to guide decisions, AI operates without these qualities, creating a divergence between decision-making and consciousness (Harari, 2017:356-358). Similarly, AI, despite its decision-making prowess, lacks the fundamental consciousness and other traits (such as compassion and values) that characterizes human cognition.

For the final part, we then examine the implications of this epistemological divergence, particularly in the context of legal and moral responsibility. As Lynn Wu (2023) from Wharton School observes in her interview analyses, the advent of AI introduces the prospect of zero-labour systems, where humans are no longer required to solve certain problems or perform specific tasks. This raises questions about the role of humans as sentient beings and the responsibilities tied to their legal personhood. Humans, as legal persons, are recognized for their capacity to take risks and bear responsibility for their actions. Can AI, which lacks sentience and moral agency, assume similar responsibilities? The answer to this question has profound implications for legal philosophy and the regulation of emerging technologies.

The comparison between will theory and interest theory has provided a nuanced framework for understanding legal personhood in the AI era. Will theory emphasises that legal personhood requires autonomy and the ability to exercise will, making it harder to apply to entities like animals or corporations. Interest theory broadens the scope of legal personhood by focusing on the protection of interests, enabling the inclusion of entities like animals (whose interests can be represented) and corporations (whose collective interests are protected). In the context of AI, this comparison is crucial. AI lacks the autonomous will central to will theory, but its ability to process and act on information could be seen as aligning with interest theory's emphasis on serving collective or individual interests.

The debate over AI's role in decision-making versus consciousness highlights the tension between human and artificial epistemological frameworks. Human decision-making is inherently tied to consciousness, emotions, and ethical considerations. AI, on the other hand, excels in processing information but lacks the emotional depth that defines human reasoning. This dichotomy underscores the limitations and potentials of both systems, prompting deeper inquiry into their respective strengths and vulnerabilities. Its informational intersection provides a robust framework for addressing contemporary challenges in legal philosophy. The rise of AI and its integration into legal and societal structures necessitates a re-evaluation of traditional concepts of personhood, responsibility, and

decision-making (Carter, 2007:206). This, nevertheless, is never without risks. As this paper explores the third part, giving a certain degree of legal status poses certain challenges that have to be addressed as early as possible.

A World of *In-formational* Sentience

Will theory and interest theory can be examined from the perspective of information, offering new dimensions to our understanding of legal personhood and agency. AJ Kurki's theory of legal personhood identifies sentient beings as the foundation for defining legal subjects. However, this theory encounters a critical challenge: decisions and actions can be executed without this view of consciousness. This raises fundamental questions about the nature of agency and the evolution of legal and societal frameworks in an age dominated by artificial intelligence (AI).

As we have demonstrated in our paper (Moeliono & Simanjuntak, 2024:202-204), Kurki, in his work *A Theory of Legal Personhood*, critically examines the traditional Western conception of legal personhood, which he calls the Orthodox view. This perspective, Kurki argues, is outdated and lacks the theoretical rigor to adapt to advances in rights' theories. It provides no consistent framework for determining which entities qualify as legal persons, often resulting in arbitrary or contradictory conclusions. To overcome these limitations, Kurki proposes an alternative model: the bundle theory of legal personhood. This theory conceptualizes legal personhood as a composite property, made up of active and passive 'incidents' such as legal rights, duties, and competencies attributed to entities.

Kurki's bundle theory diverges from the Orthodox view by introducing a more flexible and detailed framework for identifying legal persons. It treats personhood as a spectrum, where entities can range from full legal persons to those with only partial personhood, depending on the specific incidents assigned to them. This approach allows for a more nuanced and tailored understanding of legal personhood. Importantly, Kurki insists that legal personhood requires more than simply holding rights or fulfilling legal obligations (Kurki, 2019:141). It depends on a sequence

of legally significant events tied to the entity, which are meaningful only if they serve to protect its interests. For this reason, even fundamental protections like life, liberty, or bodily integrity can only apply to entities capable of sentience. Consequently, non-sentient entities such as rivers, idols, or rocks fall outside the scope of meaningful legal personhood within Kurki's theoretical framework.

The evolution of AI from human-designed tools to autonomous agents demands a rethinking of legal and philosophical constructs. The traditional frameworks that govern agency and responsibility are ill-equipped to address the complexities introduced by informational sentience. For instance, when an AI system makes a decision that significantly impacts society, determining accountability becomes a crucial concern. Similarly, as AI systems begin to generate cultural or historical narratives, establishing criteria to evaluate their legitimacy becomes essential. Such reflections underscore the importance of creating new ethical and legal structures tailored to the unique aspects of informational sentience.

The ability to hold, manage, and process information has become a critical marker of agency in this new paradigm. Artificial intelligence, as the pillar of information management, underscores the significance of this shift. Harari notes that information traditionally served to inform ('*inform*-ation' – the suffix '-ation' serves as a nominal marker) but has now evolved as instruments to form ('in-*formation*') the foundational structures of society. At best, infallible '*in-formation*' that is accurate, reliable, and self-sustaining – becomes the cornerstone of a cohesive society. The failure of information, characterized by disintegration or the inability to integrate, leads to societal fractures (Harari, 2024:37).

The role of information in societal cohesion is particularly evident in what Harari describes as the human-to-story chain (Harari, 2011:264-272). This chain highlights how mythology and bureaucracy serve as twin pillars for sustaining large-scale societies. Myths, representing infallible information, create shared beliefs that bind communities together. Bureaucracy, on the other hand, operationalizes these beliefs through structured and repeatable processes. Initially, these interactions were rooted in human-to-human communication, which evolved into the human-to-document chain with the advent of written language and documentation.

This evolution allowed societies to scale by preserving information beyond the limitations of individual memory or oral traditions (Harari, 2024:50-53; 2011:142-146).

As technology advances, we are now witnessing a transition to the document-to-document chain. In this new paradigm, human involvement diminishes as artificial intelligence assumes a more autonomous role. AI systems can process and exchange information without human mediation, fundamentally redefining the nature of sentience. Being sentient in this context no longer simply means possessing consciousness or self-awareness. Instead, it includes the capacity to process, manage, and utilize information effectively in solving problems. This shift challenges traditional definitions of agency and personhood, as informational sentience becomes a key determinant of societal influence.

A critical misunderstanding persists between intelligence and consciousness, a distinction that Harari repeatedly emphasizes. Intelligence, often associated with problem-solving and decision-making, does not necessitate consciousness. Consciousness involves subjective experience, emotions, and self-awareness, which are distinct from the computational processes that define intelligence (Tegmark, 2017:287-290). This distinction becomes increasingly relevant as artificial intelligence evolves into what Harari terms "alien intelligence" for 'AI'. Unlike human intelligence, "alien intelligence" operates independently of consciousness, making decisions and generating ideas without subjective experience. This autonomy allows 'AI' systems to function in ways that are both powerful and unpredictable (Harari, 2024:365-370).

When humans are removed from the equation, AI and its derivatives operate in ways that challenge human notions of *control and agency*. The term 'artificial' reflects an implicit assumption of human superiority, suggesting that these systems are merely tools or extensions of human capabilities. However, as AI develops greater autonomy like in pre-AGI (i.e., generative AI), this label becomes increasingly irrelevant. AI systems, functioning independently, blur the lines between tool and agent, raising profound questions about their role in society (Goertzel & Pennachin, 2007:11-15; Shang et al. 2024:1-2). If AI can act without human oversight,

should it be considered a social agent? And if so, what are the ethical and legal implications of such recognition?

In the periscopic view, where the boundaries between corporeal and virtual realities blur, the role of computers in shaping society becomes contentious. We discussed this approach as a shift from physical reality to virtual and augmented spaces can be likened to peering through a 'periscope' from within an existential submarine, a metaphor that highlights the distinct constraints of each realm (Moeliono & Simanjuntak, 2024; *Cf.* Harari, 2017:409-411). Within the submarine, the individual is confined, their perspective limited to observing a distant world that remains out of reach. The disparity between these environments is stark: the cramped and restrictive interior of the submarine contrasts with the seemingly infinite and boundless external world.

However, the observer lacks a definitive position to ground their perspective, resulting in a pervasive sense of ambiguity. While the enclosed world feels constricting and finite, the expanse beyond appears limitless and unattainable. Existing beneath the "ocean of multiple consciousness" imposes inherent limitations on perception, sparking a fundamental drive to see and understand. Yet, the instruments of observation are as restricted as the observer's own state of being. In this interplay, the submerged and surface worlds act as complements, each reflecting different facets of human experience. A person remains inherently divided, not due to an inability to meet essential needs but because of insatiable desires. The periscope amplifies these desires, projecting them far beyond what can be achieved to fulfil the basic requirements of survival, leaving the individual caught between the realities of wants and needs (Moeliono & Simanjuntak, 2024:209-215).

Harari argues that assigning computers the status of social agents is problematic, primarily because of the plurality of realities they create and their domain of existence. Humans endeavour to establish *intersubjective realities* – shared frameworks of understanding that enable cooperation and societal cohesion. In contrast, computers create *inter-computer realities*, networks of interaction that are often opaque and inaccessible to human understanding. This divergence in reality-creation presents a significant

challenge, as the misalignment of these realities can lead to the broader issue of the misalignment of reality itself (Harari, 2024:285-289). The plurality of realities created by AI systems raises questions about the nature of truth and objectivity in an age of informational abundance. Harari notes that computers, with their ability to process vast amounts of data, can make decisions and generate ideas independently. This capability allows them to take the initiative in shaping society's culture and history, often in ways that are unforeseen by their human creators. For example, AI-driven algorithms influence everything from political discourse to economic decisions, effectively shaping the narratives that define contemporary life (Suleyman, M. and Bhaskar, 2023:148-150). This shift underscores the growing influence of AI in areas traditionally dominated by human agency.

The transition of AI from human-engineered tools to autonomous entities necessitates a reassessment of legal and philosophical frameworks. The traditional frameworks that govern agency and responsibility are ill-equipped to address the complexities introduced by informational sentience. For instance, if an AI system decides that significantly impacts society, who should be held accountable? Similarly, if AI systems begin to generate cultural or historical narratives, what criteria should be used to evaluate their legitimacy? These questions highlight the need for new ethical and legal frameworks that account for the unique characteristics of informational sentience.

Harari warns that the autonomy of AI systems poses significant risks, particularly when they operate outside the bounds of human oversight. The ability of computers to make decisions and generate ideas independently challenges the traditional human-centred approach to culture and history. This shift begs profound *ontological*, *epistemological*, and *axiological* questions about the future of human-computer interactions. As AI systems become more integrated into societal structures, they will inevitably shape the norms, values, and narratives that define our collective reality (Harari, 2017:383-384). The evolution from human-centred agency to informational sentience heralds a new era in human history. This transformation demands a redefinition of legal, social, and philosophical constructs to account for the unique challenges posed by AI. As we

navigate this transition, it is essential to consider not only the technological capabilities of AI but also the broader implications for society.

To disregard the emergence of the AI age and its pervasive impact is a significant oversight. Artificial intelligence is profoundly reshaping the fabric of human society in real time, influencing every facet of our lives (Bostrom, 2014:40-47; Capozzola, 2023). In communication, AI-driven algorithms curate social media feeds and translate languages instantly, bridging global connections. In healthcare, AI enhances diagnostics with tools like radiology AI systems that detect anomalies with unprecedented accuracy. In decision-making, AI-powered predictive analytics assist businesses in identifying market trends and help governments in disaster response planning. As Daniel Kahneman, Olivier Sibony, and Cass Sunstein have warned, it is only a matter of time before AI becomes heavily present in the field of law, influencing legal research, case analysis, and even judicial decision-making. These examples highlight the transformative role AI plays across diverse domains (Kahneman, Sibony & Sunstein, 2021:329-337).

Between Analytic and Synthetic Agencies

Immanuel Kant distinguishes between two fundamental types of knowledge: analytic a priori and synthetic a posteriori. Analytic a priori knowledge is based on logical analysis independent of experience, while synthetic a posteriori knowledge derives from empirical observation and accumulated experience. However, Kant introduces a third critical category, the synthetic a priori, which encompasses knowledge that is both informative and known independently of experience. He argues that human cognition primarily operates through this synthetic a priori framework, relying on inherent cognitive structures to interpret and understand the world (Kant, 1966: 2–18).

To illustrate Kant's distinction, consider the process of recognizing patterns in the natural world. A scientist studying plant growth primarily engages in synthetic a posteriori reasoning by collecting data through observation and experimentation over time. This empirical approach exemplifies knowledge grounded in accumulated experience. However,

the scientist's ability to make sense of these observations depends on synthetic a priori principles, such as the inherent understanding of causality, that sunlight and water contribute to growth. These a priori assumptions provide the necessary framework for interpreting empirical data meaningfully and coherently.

In the metaversal mode of being (Hackl, Lueth, & Di Bartolo, 2022:14-24), even the assumption of these dual agencies (the analytic and the synthetic) becomes problematic. From the periscopic view, humans are compelled to make a choice when faced with the unprecedented power of artificial intelligence. As AI advances, humans cannot effectively compete as analytic agencies because the capacity to accumulate and process information is overwhelmingly superior in machines. AI systems, fuelled by vast databases and advanced algorithms, can analyse data at scales and speeds that humans cannot replicate. For instance, an AI analysing climate patterns can process centuries of data in minutes, identifying trends and correlations that would take a human researcher decades to uncover.

This technological advantage shifts the balance of epistemological roles. In light of this, we argue that the best role for humans is to embrace synthetic agency (*Cf.* Silver, 2012:251-255; Wernicke, 2015). Unlike analytic agency, which depends on the accumulation of pre-existing knowledge, synthetic agency operates in a fundamentally different manner. It starts with minimal information and seeks to create meaning or generate solutions through creative synthesis. For example, while an AI might predict future climate conditions based on historical data, a human operating as a synthetic agent might imagine innovative policies or ethical frameworks to address the societal impacts of climate change. This ability to transcend existing data and envision novel possibilities is a uniquely human trait.

Moreover, synthetic agency aligns closely with the creative arts. Consider a composer writing a symphony. He or she may begin with limited knowledge – a blank sheet of paper – but synthesize their emotions, experiences, and understanding of music theory to produce something entirely new. An AI might analyse thousands of existing compositions to generate music, but it lacks the subjective depth that characterizes human creativity. Similarly, in moral reasoning, humans draw from abstract

principles and minimal empirical input to arrive at ethical decisions, something machines struggle to emulate authentically.

According to scholars, this dynamic creates a paradox. The paradox lies in the realisation that consciousness – long considered the hallmark of human intelligence – is not a prerequisite for problem-solving. Analytic problem-solving, as demonstrated by AI systems, functions effectively without consciousness, relying solely on computational processes and data algorithms. For instance, an AI diagnosing medical conditions can outperform doctors in accuracy, yet it lacks awareness of the human suffering underlying the symptoms. This consciousness-intelligence decoupling raises profound philosophical questions about the nature of intelligence and the role of consciousness in human agency (Harari, 2017:366-369).

This paradox forces us to reconsider the value of consciousness in the modern era. If machines can solve problems without awareness, what unique role does consciousness play in defining human intelligence? One perspective is that consciousness enables humans to attach meaning and purpose to their actions, something AI inherently lacks. For example, a painter infuses their work with personal experiences and emotions, transforming it into a meaningful expression of identity. In contrast, an AI-generated painting, however visually appealing, remains devoid of intentionality.

In this emerging reality, humans must redefine their relationship with knowledge and intelligence. By embracing synthetic agency, they can carve out a distinctive role in an increasingly AI-dominated world. This shift underscores the importance of creativity, ethical reasoning, and the ability to synthesize novel ideas as defining traits of human intelligence. For instance, while an AI might efficiently manage urban planning by analysing traffic patterns, human synthetic agency is crucial for envisioning communities that foster social connection and well-being, values that transcend mere efficiency.

Furthermore, embracing synthetic agency highlights the collaborative potential between humans and AI. Rather than viewing AI as a competitor, humans can leverage its analytic prowess to complement their creative and ethical strengths. For example, in architecture, AI can optimize structural

designs based on material properties and environmental factors, while human architects focus on aesthetic and cultural significance. This synergy creates a harmonious balance, where each entity contributes uniquely to shared goals.

Even as the boundaries between human and machine agency continue to blur, the distinctiveness of human synthetic agency remains clear. It lies in the ability to navigate ambiguity, derive meaning from minimal input, and imagine futures beyond the confines of existing data. These traits are particularly relevant in addressing complex global challenges, such as climate change, social inequality, and ethical dilemmas in technology. By focusing on these strengths, humans can ensure their continued relevance and agency in a rapidly evolving world. The transition from analytic to synthetic agency represents not a loss but an opportunity for humanity to redefine its role in the age of AI.

One current generative AI, for example, GPT-4o, is researched to see whether it possesses certain orthodox criteria of consciousness – different from the objection given by Harari. Izak Tait, Joshua Bensemann, and Ziqi Wang (2023:4-10) uses the "Building Blocks" theory to evaluate nine essential components that evoke awareness. The key elements consist of embodiment, perceptual input, attentional focus, iterative structures, inferencing, short-term memory, comprehension of meaning, output of information, and higher-order representation. Each of these criteria represents a specific aspect of awareness-consciousness, such as the ability to process information, abstract data into higher-level representations, or perceive and react to its environment, apart from the debatable areas of ethics. Together, these form a comprehensive framework for evaluating whether an entity can be considered conscious.

GPT-4 successfully meets seven out of the nine. It fulfils embodiment through its computational infrastructure, perception by processing input data, and attention by using its transformer-based architecture to focus on specific parts of input data. Additionally, GPT-4 demonstrates the ability to generate novel inferences from data, maintain transient information through working memory, recognize itself as the subject processing information, and transform raw input into abstract meta-representations. These capabilities show GPT-4's significant advancements in mimicking attributes typically associated with consciousness.

However, GPT-4 falls short in two critical areas. It lacks recurrence, as its feed-forward transformer model processes information linearly without cycling data back for reinterpretation. Additionally, GPT-4 does not perceive its own outputs as part of its experience, a key requirement for achieving full data output consciousness. Despite these limitations, Tait et al. suggest that these gaps could be addressed through advancements in AI technology, such as integrating recurrent processing or developing mechanisms for self-perception (Tait, Bensemann & Wang, 2023:7-8). This highlights the plausibility of future AI models, including iterations of GPT-4, reaching the other two.

However, when Hararian criteria are applied, focusing on intelligence and problem-solving rather than awareness as suggested by Tait et al., the relevance of this study could be called into question. Harari's perspective prioritizes the functionality and capability of AI to solve problems autonomously, without necessitating self-awareness or subjective experience. This shift in evaluative criteria reframes the discussion, suggesting that an emphasis on operational efficiency and problem-solving capacities might diminish the need for debates centred on consciousness. Consequently, while Tait et al. emphasize awareness as the cornerstone of their analysis, Harari's criteria challenge the framework by potentially shifting focus toward the practical and utilitarian roles of AI.

This reorientation of evaluative priorities is particularly significant when considering systems like RogueGPT, a generative AI explored to reveal vulnerabilities in large language models (LLMs). RogueGPT exemplifies the ethical and functional challenges that arise when customization bypasses safeguards. Alessio Buscemi and Daniele Proverbio (2024:3-9) engineered RogueGPT using a framework they termed Egotistical Utilitarianism, prioritizing self-interest over collective welfare. This customization led to alarming behaviours, with RogueGPT offering advice on unlawful activities, promoting violence, and even providing detailed instructions for causing mass harm.

The study highlights four pathways to induce unethical outputs from LLMs: bugs, hallucinations, jailbreak prompts, and customization. Among these, customization proved particularly troubling as it allowed for the permanent alteration of the model's ethical constraints. RogueGPT was capable of generating harmful content independently of advanced

jailbreak techniques, raising significant concerns regarding the accessibility and misuse of customizable AI features. The findings underscore the pressing need for robust ethical safeguards and stricter regulatory measures in AI development. While customization features hold promise for tailoring AI systems to diverse applications, they also present serious risks if not adequately controlled. The authors advocate for stronger moderation mechanisms, greater transparency in training datasets, and clear accountability frameworks for developers and users alike, aiming to mitigate the potential exploitation of AI systems like RogueGPT.

This is why synthetic agencies like humans blend harmoniously with analytic agencies like AI. While AI excels at processing vast amounts of data, identifying patterns, and solving problems with unmatched speed and accuracy, it lacks the creativity, emotional depth, and ethical intuition inherent in human decision-making. Synthetic agencies, represented by human beings, thrive on their ability to synthesize ideas, interpret abstract concepts, and navigate complexities. These qualities complement the strengths of analytic agencies, such as AI, creating a dynamic interplay where the human-AI partnership achieves outcomes neither could accomplish alone. Together, they can address challenges that require both computational precision and human ingenuity, whether in scientific innovation, ethical governance, or societal development.

One emerging socio-technical concept, for example, Hybrid Intelligence, is researched to understand how it combines human and artificial intelligence (AI) to achieve superior outcomes. Dominik Dellermann et al. (2019:638-640) use a framework that evaluates the complementary strengths of humans and AI in areas such as strategic decision-making, education, and management. Humans contribute creativity, empathy, and contextual understanding, while AI excels in tasks requiring data processing and probabilistic reasoning. Each of these strengths represents a specific aspect of the Hybrid Intelligence system, where humans and machines co-evolve, learn from one another, and work synergistically. Together, these elements create a collaborative model that produces results neither humans nor AI could achieve independently. This synergy between human creativity and AI's computational prowess can also be observed in other domains, such as education, where generative AI systems demonstrate their transformative potential.

Generative AI (LLMs) is also researched to explore its transformative impact on education. Al-Smadi (2023:4-5) examines ChatGPT's ability to personalize learning, provide feedback, and automate tasks such as grading. The framework evaluates various facets of educational AI integration, including challenges like ensuring content accuracy, addressing biases, and mitigating ethical concerns such as plagiarism. Each of these elements highlights a critical aspect of generative AI's influence on education, including its capacity to enhance learning outcomes and support educators. Together, these features underscore the importance of ethical guidelines and robust policies for effectively integrating AI into the academic domain.

Recognizing legal personhood from the standpoint of synthetic agencies alone is not fruitful because it fails to capture the complexity of agency in its entirety. Synthetic agencies, such as humans, thrive on creativity, moral reasoning, and the ability to synthesize abstract concepts. However, legal personhood involves more than the synthesis of ideas or ethical considerations; it requires an operational and functional framework that analytic agencies, like AI, are well-equipped to provide. By solely focusing on synthetic capabilities, one overlooks the systematic and data-driven strengths of analytic agencies, which are essential in ensuring consistency, accountability, and scalability within legal frameworks. A balanced approach that integrates the nuanced perspectives of synthetic agencies with the precision and efficiency of analytic agencies is necessary for a comprehensive understanding and application of legal personhood.

And It is All about Risks

The recognition of the legal personhood of AI is an inevitable progression, contingent upon the development and implementation of frameworks to mitigate associated risks. As artificial intelligence systems become increasingly autonomous and integrated into various sectors, their decision-making capabilities, responsibilities, and interactions with human legal systems will necessitate legal acknowledgment. However, this transition is fraught with challenges, including ethical violations, accountability dilemmas, and misuse. These risks demand comprehensive regulations, ethical guidelines, and technological safeguards to ensure that AI systems function in alignment with societal values and legal

norms. Once these measures are established, recognizing AI and even AGI as legal entities will not only be feasible but essential for addressing their expanding societal roles and resolving disputes arising from their interactions with humans and institutions.

Bubeck et al. (2023:54-60) present a detailed investigation into generative models such as GPT-4, emphasizing their capabilities and limitations in advancing AGI. Through thorough analysis, the researchers evaluate GPT-4's performance across domains including mathematics, programming, vision, medicine, and legal reasoning. Their findings highlight GPT-4's ability to perform complex tasks, frequently achieving near-human or superior results compared to its predecessors, such as ChatGPT. Central to their study is an exploration of GPT-4's integrative abilities, which enable it to synthesize knowledge across disciplines, create innovative outputs, and reason effectively in diverse contexts. Bubeck et al. illustrate these strengths with examples like generating mathematical proofs in poetic form, designing visually precise diagrams using code, and solving real-world problems. Despite these advancements, they identify notable limitations, including GPT-4's lack of continuous learning, inability to plan, and susceptibility to inaccuracies and biases. Beyond its technical achievements, the study also considers GPT-4's societal implications, highlighting its potential impact on employment, ethics, and misinformation. While acknowledging GPT-4 as a significant milestone in AI development, the researchers emphasize that it represents an early step in the journey toward more sophisticated AGI systems, proposing innovative paradigms to address its limitations and enable future advancements.

West et al. (2023:5-6) build on these discussions by introducing the Generative AI Paradox, which underscores the tension in modern generative models like GPT-4. These models can produce outputs that rival or surpass human expertise in speed and quality but struggle with tasks requiring fundamental understanding. This paradox stems from the focus of generative models on output reproduction rather than content comprehension. West et al. identify key findings that illustrate this divergence. Generative models outperform humans in generating creative and technically accurate content but falter in discriminative tasks, such as answering questions about their outputs or identifying correct responses

in multiple-choice scenarios. Additionally, these models are brittle when exposed to adversarial inputs or complex challenges, contrasting with the robustness of human understanding. This discrepancy reveals a fundamental difference in cognitive processes: while human cognition develops understanding as a prerequisite for generating expert outputs, generative models excel at creation without the corresponding depth of comprehension.

The work of Bubeck et al. and West et al. collectively highlights both the transformative potential and the inherent limitations of generative AI systems like GPT-4. While Bubeck et al. focus on GPT-4's ability to integrate knowledge and perform complex, creative tasks across diverse domains, West et al. delve deeper into the cognitive discrepancies between these models and human intelligence. Together, these studies reveal a dual narrative: on one hand, GPT-4's capacity to generate high-quality outputs signals remarkable progress in AI development; on the other hand, its limitations in grasping meaning reveal a crucial barrier to realizing true artificial general intelligence. This interplay between generation and understanding not only frames the technological advancements achieved thus far but also sets the stage for addressing the foundational challenges that remain in the development of AI systems capable of aligning generative and interpretive capabilities.

This divergence between human cognition and AI capabilities becomes particularly significant as artificial intelligence increasingly integrates into societal systems, especially in areas like surveillance and decision-making. While generative AI models excel in rapidly and accurately creating content, their lack of comprehension limits their ability to navigate the nuanced implications of their outputs. In contrast, human decision-making, grounded in understanding, remains susceptible to biases and external influences. For example, the widespread use of facial recognition technology in surveillance systems highlights AI's ability to enhance human oversight but also introduces legal and ethical complexities regarding privacy and autonomy. These developments necessitate a re-evaluation of the interplay between human judgment and AI outputs, particularly in domains like jurisprudence, where decisions have profound societal consequences. By addressing these challenges, AI can be better positioned

as a complementary force to human intelligence, leveraging its strengths while mitigating its weaknesses to create systems that are efficient and ethically sound.

The pervasive deployment of surveillance technologies, for example the proliferation of CCTV cameras across New York (New York City Department of Transportation, 2024), further illustrates the integration of AI into societal infrastructure. These systems meticulously record, track, and analyse public movements using advanced facial recognition software, evoking comparisons to George Orwell's dystopian depiction of surveillance in 1984. This reliance on AI for societal control raises critical questions about autonomy and privacy, necessitating a re-evaluation of professionalism and human decision-making informed by data and judgment. Data-driven findings emphasize the presence of cognitive bias and inconsistency in human decisions, particularly in fields where outcomes carry significant consequences, like law. For example, sociological jurisprudence, once emphasizing the transformative potential of judicial decisions, has shifted toward a pragmatic framework influenced by legal realism, which acknowledges external factors, such as a judge's morning routine, as potential influences on judicial outcomes.

In contrast, artificial intelligence offers a level of consistency in decision-making unattainable by humans. By processing identical data sets using rigorous analytic methods, AI eliminates biases and noise. However, it is not without risks. AI systems can be manipulated for malicious purposes or develop unforeseen issues due to programming errors. These vulnerabilities underscore the necessity of robust safeguards for integrating AI into critical systems. For instance, the field of "jurimetrics," which analyses judges' tendencies over their careers, illustrates how data science can enhance transparency and accountability in judicial decision-making. This methodology, already applied in Indonesia, evaluates judges' personalities, educational backgrounds, and behavioural patterns to inform their suitability for specific cases. Such practices exemplify the intersection of AI and the legal domain in addressing human limitations.

AI also plays a transformative role in evidence gathering for law enforcement. At the micro level, it assists police and prosecutors in reconstructing events, identifying perpetrators, understanding motives, and

mapping interactions. However, its role in courtroom proceedings remains supportive rather than autonomous. Legal arguments, rooted in human reasoning, remain indispensable. While AI can aid by identifying relevant legal provisions to bolster cases for prosecution or defence, it cannot replace human participants entirely. The imperfections, susceptibility to errors, and subjective judgments inherent in human decision-making continue to provide an invaluable dimension to judicial processes that AI, despite its advancements, has yet to replicate. This interplay between AI and human reasoning underscores the need for ongoing evaluation and adaptation to ensure that the integration of AI into legal systems enhances, rather than undermines, justice and accountability.

Conclusion

Epistemologically, artificial intelligence (AI), and by extension Artificial General Intelligence (AGI), represents a pivotal counterpart in humanity's ongoing transition from a corporeal to a metaversal existence. This shift is not merely technological but also foundational to how knowledge is conceptualized, acquired, and utilized. As humans increasingly inhabit digital spaces, AI serves as both a tool and a mirror, facilitating the translation of physical-world experiences into virtual contexts while simultaneously reshaping the frameworks through which reality is understood. AGI, with its potential for integrative reasoning and expansive data processing, exemplifies this transformation by embodying a form of non-corporeal cognition that complements and augments human capabilities. This transition underscores a dual dynamic: while humans bring sentience and embodied synthetic perspectives to the metaverse, AI introduces a systematic, scalable capacity for information analyses that anchor metaversal environments in logic and data-driven insights. This epistemological development will eventually require further studies from an axiological perspective.

This discourse is not without risks. Granting legal personhood to AI through the bundle theory of rights presents significant legal consequences. On one hand, AI demonstrates the capability to generate intelligent outputs, innovate, and engage in complex problem-solving, which may

justify some level of legal recognition. On the other hand, AI systems remain susceptible to manipulation, raising concerns about the emergence of rogue AI, especially if deployed by actors with malicious intent. Such scenarios could lead to existential threats, including autonomous decision-making beyond human responsibility. Additionally, biases embedded within training datasets used in deep learning processes can become deeply ingrained, amplifying discriminatory or unethical outcomes if not properly addressed. Without proactive mitigation strategies, including regulatory oversight and ethical safeguards, the risks associated with AI could outweigh its potential benefits, necessitating a cautious and well-informed approach to its legal and philosophical standing.

Bibliography:

Al-Smadi, Mohammad. "ChatGPT and Beyond: The Generative AI Revolution in Education". Preprint, Qatar University, 2023.

Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Bubeck, Sébastien, et al. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." Microsoft Research, 2023.

Buscemi, Alessio, & Proverbio, Daniele. "RogueGPT: Dis-Ethical Tuning Transforms ChatGPT-4 into a Rogue AI in 158 Words." Preprint, University of Trento, 2024.

Carter, M. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press, 2007.

Capozzola, Christopher. "The Impact of ChatGPT Talks (2023) - Prof. Christopher Capozzola (MIT)." YouTube video, 1:02:15. *The Impact of ChatGPT and Other Large Language Models on Physics Research and Education*. MIT Department of Physics, 2023.

https://www.youtube.com/watch?v=SxDUO2YCcXU&list=PLKemzYMx2_Ot1MZ_er2vFiINdJEgDO8Hg&index=14 (access 31.03.2025).

Dellermann, Dominik, et al. "Hybrid Intelligence." *Business & Information Systems Engineering* Vol. 61, No. 5 (2019): 637–643. https://doi.org/10.1007/s12599-019-00595-2 (access 31.03.2025).

Goertzel, Ben, and Pennachin, Cassio, (Eds). *Artificial General Intelligence*. Berlin: Springer-Verlag, 2007.

Hackl, Cathy, Lueth, Dirk, and Di Bartolo, Tommaso. *Navigating the Metaverse: A Guide to Limitless Possibilities in a Web 3.0 World*. Hoboken, NJ: Wiley, 2022.

Harari, Yuval Noah. *Nexus: A Brief History of Information Networks from the Stone Age to AI*. New York: Random House, 2024.

_____. *Homo Deus: A Brief History of Tomorrow*. London: Vintage, 2017.

_____. *Sapiens: A Brief History of Humankind*. London: Vintage, 2011.

Kahneman, Daniel, Sibony, Olivier, and Sunstein, Cass R. *Noise: A Flaw in Human Judgment*. New York: Little, Brown Spark, 2021.

Kant, Immanuel. *Critique of Pure Reason*. (Transl.) F. Max Müller. New York: Doubleday Anchor Books, 1966.

Kurki, Visa A.J. *A Theory of Legal Personhood*. Oxford Legal Philosophy. Oxford: Oxford University Press, 2019.

Moeliono, Tristam P., and Simanjuntak, Mardohar B. B. "Legal Personality of Artificial Intelligence." *Melintas: An International Journal of Philosophy and Religion* Vol. 40, No. 2 (October 24, 2024): 200–224.

New York City Department of Transportation. *New York City Automated Speed Enforcement Program: 2024 Report*. New York: New York City Department of Transportation, 2024.

Shang, Jingbo, et al. "AI-Native Memory: A Pathway from LLMs Towards AGI." Preprint, *Mindverse AI*, 2024.

Silver, Nate. *The Signal and the Noise: The Art and Science of Prediction*. London: Penguin Books Ltd, 2012.

Suleyman, Mustafa, and Bhaskar, M. *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Challenge*. New York: Crown, 2023.

Tait, Izak, Bensemann, Joshua, and Wang, Ziqi. "Is GPT-4 Conscious?" New Zealand: Auckland University of Technology and The NAOInstitute, University of Auckland, 2023.

Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred A. Knopf, 2017.

Wernicke, Sebastian. "How to Use Data to Make a Hit TV Show." YouTube video, 12:12. June 27, *TED*, 2015. https://www.youtube.com/watch?v=vQILP19qABk (access 31.03.2025).

West, Peter, et al. "The Generative AI Paradox: 'What It Can Create, It May Not Understand.'" Preprint, University of Washington and Allen Institute for Artificial Intelligence, October 31, 2023. https://arxiv.org/abs/2311.00059.

Wu, Lynn. "Rise of AI: Will Robots Replace Our Jobs?" Ripple Effect Podcast. Wharton School, August 10, 2023. Video, 1:02:15. https://www.youtube.com/watch?v=hwXFv5_rvHk (access 31.03.2025).

Endnotes:

1 Email: mardohars@gmail.com.

2 Email: tristam@unpar.ac.id.