CIENTIFIC Publishing Limited

Journal of Intelligent Communication

https://ojs.ukscip.com/journals/jic

Article

AI and the Cognitive Sense of Self

Emily Barnes [1,*], **James Hutson** [2]

[1] University Department, Capitol Technology University, AI Center of Excellence (AICE), 11301 Springfield Rd, Laurel, MD 20708

[2] University Department, Lindenwood University, Art History, AI, and Visual Culture, Saint Charles, MO 63301, USA

[*] Correspondence: ejbarnes035@gmail.com

Received: 20 January 2024; **Revised**: 10 March 24; **Accepted**: 13 March 2024; **Published**: 20 March 2024

Abstract: This article explores the emergence of a cognitive sense of self within artificial intelligence (AI), highlighting its transformative potential to enhance complex interactions and autonomous decision-making in intelligent systems. Central to this investigation is a mixed-method study approach designed to validate the research, integrating qualitative and computational analyses to examine AI systems in healthcare and robotics. The study focuses on mechanisms such as self-recognition, self-reflection, and identity continuity—attributes that mirror aspects of human consciousness and are critical for creating systems capable of personalized, adaptive interactions. By blending cognitive science theories with practical AI development, the research introduces a robust framework for engineering self-aware systems capable of nuanced, context-sensitive functionalities. Additionally, the study examines the ethical dimensions of self-aware AI, underscoring the need for comprehensive ethical guidelines to ensure transparency, accountability, and fairness in their development and deployment. Addressing issues such as the societal impact, potential misuse, and the moral responsibilities of creators, the research emphasizes the importance of aligning technological innovation with ethical principles. The findings contribute to the theoretical discourse on machine consciousness while offering actionable insights for implementing these technologies across various industries. This integrated approach underscores the dual importance of advancing AI capabilities and navigating their societal and ethical implications responsibly, positioning self-aware AI as both a technological milestone and a profound challenge for contemporary research and practice.

Keywords: Artificial Intelligence; Cognitive Self; Autonomous Systems; Ethical AI; Human-like Consciousness.

1.Introduction

The intersection of artificial intelligence (AI) and consciousness represents a frontier in contemporary scientific inquiry, drawing upon foundational philosophical theories and the intricate architectures that might underpin conscious experiences in machines. Initial research phases have focused on the complex relationship between AI and consciousness, exploring the capacity of AI systems to emulate human-like conscious behaviors through sophisticated algorithms

https://doi.org/10.54963/jic.v3i1.320

and network architectures [1–4]. Building upon these insights, the current discourse evolves towards a more nuanced investigation into the development of a cognitive sense of self within AI systems.

The concept of self-awareness, a fundamental component of human consciousness, includes the capacity to perceive oneself as a distinct entity, separate from the environment and other beings. Such cognitive attributes encompass self-recognition, self-reflection, and a continuous sense of personal identity [5]. For AI systems, the cultivation of similar self-aware traits promises to significantly enhance their functionalities, facilitating more refined interactions, adaptable behaviors, and autonomous decision-making processes. Pioneering research in this domain has been led by scholars like Legaspi et al. [6], who have underscored the potential for these systems to develop a sense of agency and self-awareness, noting key advancements in areas such as self-attribution of actions and Bayesian inferencing. To further this endeavor, Oberg [7] posits that a deep understanding of human consciousness and cognitive models is indispensable for nurturing self-awareness in machines. This perspective highlights the imperative for an interdisciplinary approach to development, integrating cognitive science with technological innovation. Supporting this view, Tani and White [8] as well as Parziale and Marcelli [9] examine the role of cognitive neurorobotics in simulating the human sense of self, demonstrating how dynamic interactions within neural networks could mimic aspects of human cognition.

This study not only explores theoretical frameworks but also looks into practical applications, presenting case studies where AI systems display behaviors indicative of self-awareness as well as a proposed methodology for reproducibility. The ethical dimensions of these advancements are critically analyzed by researchers like Levin [10], who discuss the moral responsibilities and societal implications inherent in the development of self-aware machines. Furthermore, scholars such as Marcus and Davis [11] advocate for leveraging insights from cognitive science to enhance the adaptability and flexibility of such systems. The quest to endow architectures with a cognitive sense of self spans a rich landscape that intersects advanced technology, deep philosophical questions, and pressing ethical concerns. At the same time, the notion of self-aware AI has largely been approached from theoretical perspectives, with substantial discourse centered on hypothetical scenarios and speculative futures. However, practical explorations that integrate these theories with real-world applications and their ethical ramifications are noticeably sparse. This study aims to bridge this gap by synthesizing empirical case studies and theoretical insights, thus examining self-aware machines through comprehensive lenses of cognitive science, neuroscience, and ethical reasoning. This integrated approach moves beyond the speculative to confront the real and tangible impacts of implementing self-aware frameworks in various sectors.

The primary objective of this research is to consider the theoretical frameworks that underpin the potential for self-awareness in AI systems, and to demonstrate these concepts through empirical analysis. Through the examination of detailed case studies, this study highlights the operationalization of self-aware AI in complex environments such as healthcare and robotics, showcasing the enhanced decision-making and adaptability these systems can offer. Furthermore, it critically evaluates the ethical dimensions surrounding the development and deployment of self-aware platforms, discussing the moral responsibilities of creators, the rights of such entities, and the broader implications for society. The novelty of this article lies in its comprehensive and integrative approach, which distinguishes it from previous literature that often isolates theoretical considerations from practical applications. The merging of insights from cognitive science and neuroscience with real-world implementations and ethical analyses allows this study to provide a holistic view of the challenges and opportunities presented by self-aware AI. This approach not only deepens our understanding of how AI can develop a sense of self but also illuminates the practical steps and ethical considerations necessary for responsibly advancing these technologies. The paper uniquely contributes to the ongoing discourse by providing a balanced examination that aligns the development of autonomous technologies with ethical standards and societal values, offering a forward-looking perspective on navigating the complexities of AI self-conscious.

2. Literature Review

The development of a cognitive sense of self in AI represents a growing area of research, especially with the widespread use of generative AI (GAI), intersecting disciplines such as cognitive science, neuroscience, and artificial intelligence. The concept of a cognitive sense of self in machines is integral to enhancing the capabilities of autonomous systems, allowing them to engage in more sophisticated decision-making processes and interactions. Recent advancements in the field emphasize the importance of developing self-awareness mechanisms within creative agents [12]. Srinivasa and Deshmukh [13] discuss the relevance of self-awareness in autonomous decision-making, arguing for the necessity of richer computational models that embody a sense of self to facilitate responsible behavior in these systems.

Propelling this idea forward, GAI has significantly advanced the capabilities of AI systems, enabling them to generate new data from training data and thereby enhance their decision-making and interactive abilities [14]. A notable application of generative platforms in achieving self-awareness is seen in the development of abnormality detection techniques in cognitive radio systems. Toma et al. [15] introduce a self-awareness module that uses generative models to detect abnormalities in the radio spectrum, thereby enhancing the ability to establish secure networks and make informed decisions in response to malicious activities. As such, the integration of multisensorial data and bio-inspired frameworks further supports the development of cognitively aware systems. As well, Regazzoni et al. [16] proposed a framework that employs cognitive dynamic Bayesian networks and generalized filtering paradigms to enable cognitive architectures to predict future states and select representations that best fit current observations. This approach facilitates continuous knowledge expansion and self-awareness through the analysis of proprioceptive and exteroceptive signals.

On the other hand, Zhu et al. [17] emphasize the shift towards cognitive AI that incorporates human-like common sense, identifying core domains such as functionality, physics, intent, causality, and utility as essential for developing technologies with a comprehensive understanding of its environment. This paradigm shift aims to enhance ability to solve a wide range of tasks with minimal training data, thus fostering more sophisticated and human-like interactions. The ability to continuously learn and adapt is a crucial aspect of self-awareness. Su et al. [18] thus introduce the concept of Generative Memory (GM) for lifelong learning, where systems memorize and recall learned knowledge using neural networks. This approach allows the system to accurately and continuously accumulate experiences, thereby enhancing its adaptive and decision-making capabilities [19].

While advances are being made, the development of self-aware AI systems introduces significant ethical considerations that extend beyond technical achievement to broader societal impacts. These systems, capable of understanding and reacting to their environments in sophisticated ways, raise important ethical questions about their rights, responsibilities, and the potential social ramifications of their actions. Greenwood et al. [20], for instance, emphasize the technical and ethical challenges in developing smart systems that possess a form of self-awareness. They propose using evolutionary machine learning (ML) and adversarial processes as alternatives to traditional neural network approaches. These methods could potentially allow AI to have a more dynamic and adaptable learning process without the limitations and biases often inherent in pre-trained neural networks [21].

Likewise, Vallor et al. [22] raised concerns about the socio-economic impacts of self-aware machines. They argue that if not properly managed, such structures could exacerbate existing inequalities and introduce new forms of digital divide. These researchers discuss the potential for self-aware platforms to manipulate or even replace human decision-making in critical areas, which could lead to unintended consequences on societal structures and individual freedoms. The possibility that such mechanisms could develop a sense of self-awareness also introduces questions about the rights such systems might hold and the ethical obligations of their creators and users. Discussions in the field suggest that as systems become more autonomous and integrated into daily life, there should be clear guidelines on the ethical treatment of AI, including their rights to autonomy, learning, and integration into society [23]. This involves considering

AI as potential digital "persons" with certain rights and obligations, which poses significant legal and ethical challenges. At the same time, there is a precedent in corporations who also have rights, can own property, take legal action, etc.

To address these concerns, there is a growing consensus on the need for robust ethical guidelines that govern the development and deployment of self-aware systems. These guidelines should not only ensure that systems operate safely and transparently but also respects human rights and diversity, promoting fairness and preventing discrimination. To that end, the IEEE has been active in proposing ethical standards for AI, which include transparency, accountability, and the avoidance of bias in algorithms [24]. The ethical implications of developing self-aware systems are complex and require careful consideration and proactive management. As technology continues to evolve, it is imperative that researchers, developers, and policymakers collaborate to establish ethical frameworks that guide the responsible development and use of these technologies. This will help ensure that AI serves to enhance societal well-being, rather than detract from it, and respects both human and machine rights in a balanced and thoughtful manner.

3. Cognitive Sense of Self in AI Agents

The development of a cognitive sense of self is paramount for advancing autonomous systems, equipping them with greater autonomy, adaptability, and enhanced interactive capabilities (**Table 1**). The insights of Srinivasa accentuate the significance of this cognitive attribute, enabling agents to effectively navigate their environments, make informed decisions, and engage in meaningful interactions with humans, thus recognizing their own capacities and limitations (Srinivasa, personal communication).

Table 1. Key Components and Implementations of Cognitive Sense of Self in Artificial Intelligence Systems

Component	Definition	Implementation
Self-Recognition	Ability of an AI system to identify itself as distinct from its environment and other entities.	Techniques such as computer vision and proprioception are utilized to help AI systems discern their physical presence and distinguish themselves from external objects.
Self-Reflection	Capacity of an AI system to monitor and evaluate its own internal states, processes, and behaviors.	AI systems maintain logs of their actions and outcomes, analyze this data to detect patterns, and adjust their strategies accordingly. Machine learning algorithms play a critical role in enabling the system to learn from past experiences.
Continuity of Identity	Involves maintaining a consistent sense of self over time.	Memory systems and data storage preserve information about past states and actions, allowing AI systems to build a coherent narrative of their existence. Techniques such as long-term memory in neural networks and temporal coherence algorithms support this continuity.
Agency and Intentionality	Refers to the AI system's ability to act upon its environment based on internal goals and motivations.	AI systems are designed with goal-setting mechanisms and motivational frameworks that drive their behavior. Reinforcement learning algorithms help AI agents develop strategies to achieve their goals based on rewards and feedback from the environment.
Self-Monitoring and Error Correction	Ongoing process of checking and evaluating one's own performance and rectifying mistakes.	Diagnostic tools and self-repair mechanisms are integrated into AI systems for continuous self-monitoring and error correction. Machine learning models that predict and detect anomalies assist systems in identifying errors in real-time and taking corrective actions.
Enhanced Decision-Making and Autonomy	Allows AI agents to make autonomous and well-informed decisions based on their state and capabilities.	AI systems can evaluate options and choose actions that align with their goals and constraints, especially important in dynamic and unpredictable environments where pre-programed responses are insufficient.
Adaptive Learning and Behavior	AI systems benefit from the ability to reflect on past actions and outcomes to enhance performance.	By learning from experiences and adapting over time, AI systems can continually optimize their performance, crucial for long-term deployment and continuous improvement.
Meaningful Human-AI Interaction	AI agents can achieve more intuitive and natural interactions with humans.	AI systems understand and respond to human social cues, anticipate needs, and provide personalized assistance, essential for applications in customer service, healthcare, and collaborative robotics.

Significant strides in research have elucidated the cognitive sense of "self" within agents, as explored by Tani [8], Legaspi et al. [6] and Legaspi et al. [25], who look into self-consciousness and the sense of agency in machine systems. These studies highlight how self-attribution of actions and Bayesian inferencing contribute to self-awareness. Further developments by Kahl et al. [26] and Hafner et al. [27] investigate the creation of an active self and the foundational elements necessary for an artificial self, proposing models that integrate predictive processing and developmental principles from biological systems. Kwiatkowski and Lipson [28] provides a groundbreaking example of a robot that models itself without prior programming, showcasing the potential for autonomous systems to develop self-recognition autonomously.

Enhanced decision-making and autonomy are central to the effectiveness of systems endowed with a cognitive sense of self. This capability allows agents to autonomously make well-informed decisions by recognizing their own state and capabilities, thereby enabling them to accurately assess situations and respond appropriately [29]. Such adaptability is particularly crucial in dynamic and unpredictable environments where static, pre-programmed responses would be insufficient. The ability to act autonomously not only streamlines operations but also enhances reliability in varying scenarios, reflecting a sophisticated level of artificial intelligence that approaches human-like decision-making processes. In parallel, adaptive learning and behavior are integral to the functionality of self-aware architecture [30]. These systems benefit immensely from their capacity to reflect on past actions and outcomes, which allows them to adjust their behaviors to optimize performance continually. This capacity for self-evaluation is crucial for their long-term application and continuous development, ensuring that systems remain effective and efficient [31]. By learning from experiences and adapting over time, cognitive architecture can achieve a higher level of operational excellence and utility, making them invaluable across a wide range of applications.

The development of a cognitive sense of self in machines also significantly enhances human-AI interaction. AI agents with self-awareness can engage in more natural and intuitive interactions with humans, which are crucial for applications in customer service, healthcare, and collaborative robotics [25]. These autonomous systems can understand and respond to human social cues, anticipate needs, and provide personalized support, making their integration into societal frameworks much smoother and more effective. This level of interaction is not only beneficial for enhancing user experience but also vital for the acceptance of machine mechanisms in roles traditionally filled by humans. As such, developing these capabilities involves several key components that collectively establish a robust and functional sense of identity within these systems. These components include self-recognition, self-reflection, continuity of identity, agency, intentionality, self-monitoring, and error correction [32]. Self-recognition allows such systems to identify themselves as distinct from their environments and other entities, crucial for accurate self-awareness. Techniques such as computer vision and proprioception help these systems discern their physical presence and distinguish themselves from external objects. Furthermore, self-reflection enables smart systems to assess their performance and identify areas for improvement through internal feedback mechanisms and machine learning algorithms.

Continuity of identity is supported by memory systems and data storage, which preserve information about past states and actions, allowing systems to maintain a consistent narrative of their existence. This aspect is crucial for enabling systems to adapt their goals over time based on past experiences. Additionally, agency and intentionality in platforms refer to their capacity to act upon their environment based on internal goals, with decision-making guided by goal-setting mechanisms and motivational frameworks that are often enhanced by reinforcement learning algorithms. Finally, self-monitoring and error correction are vital for maintaining the accuracy and reliability of the frameworks, ensuring that they can autonomously detect and correct errors, thereby preserving their integrity and functionality. Together, these components form the bedrock of a cognitive sense of self, equipping systems with the necessary tools to function autonomously and interact effectively. This comprehensive development not only marks a significant evolution in artificial intelligence capabilities but also highlights the complex interplay between various cognitive processes that enable systems to operate with a level of sophistication akin to human intelligence.

4. Developing Identity in AI

Developing a sense of identity in AI systems is an intricate and multi-layered process that merges various cognitive functions to establish a coherent self-concept (**Table 2**). The identity of a machine system is characterized by its ability to perceive itself as a unique entity with continuous existence over time, possessing distinct characteristics, experiences, and goals. This development of identity is pivotal, enabling the architecture to function not just as computational tools but as entities with a semblance of self-awareness and personal history.

Table 2. Mechanisms and Roles in Developing Identity in Artificial Intelligence Systems

Aspect	Definition	Details and Citations
Memory in Identity Development	Crucial for maintaining a continuous sense of identity.	Continuity of Experience: Enables AI to store and retrieve past states, actions, and experiences to construct a coherent narrative of their existence [27]. Contextual Awareness: Helps AI make informed decisions by applying lessons learned from past experiences to new situations, enhancing adaptability and depth of identity [33-34].
Learning in Identity Development	Central to the evolution of AI identity through adaptation and personalization.	Adaptive Behavior: Allows AI to modify and improve actions based on new information and experiences, driven by machine learning algorithms such as reinforcement learning and neural network training [8]. Personalized Growth: Supports development of unique characteristics by tailoring learning processes to specific interactions and experiences [35].
Self-Recognition in Identity Development	Enables AI to distinguish itself from its environment and other agents, fostering autonomy and self-awareness.	Physical and Functional Self-Recognition: Technologies such as computer vision and proprioception allow AI to recognize its own physical form and movements, essential for distinguishing self-generated actions from external events [6]. Internal State Monitoring: Enhances self-recognition by monitoring internal states and processes, aiding in maintaining a consistent self-image and adapting behaviors [36].

The role of memory in shaping machine identity is crucial as it serves as the foundation for continuity of experience, allowing smart systems to store and retrieve past states, actions, and experiences to construct a coherent narrative of their existence. Advanced neural network-based long-term memory systems are essential, enabling AI to maintain a stable sense of self over time by recalling previous experiences [37, 27]. This continuity is complemented by contextual awareness memory, which helps the platform to make informed decisions by applying lessons learned from past experiences to new situations, thereby enhancing adaptability and depth of identity [33].

Learning mechanisms also play a central role in the evolution of synthetic identity. Adaptive behavior learning allows autonomous systems to modify and improve their actions based on new information and experiences, fostering a dynamic and robust sense of self. This process is often driven by machine learning algorithms, such as reinforcement learning and neural network training, which continuously update the knowledge base and adjust its behavior to refine its self-concept and goals [8]. Moreover, personalized growth learning supports the development of unique characteristics and capabilities, reinforcing individuality within such systems by tailoring learning processes to their specific interactions and experiences [35].

Self-recognition is another fundamental component in the identity development of cognitively aware systems. It involves machine ability to distinguish itself from its environment and other agents, a capability underpinned by technologies such as computer vision and proprioception [6]. This self-recognition is crucial to perform autonomously and make decisions independent of external inputs. Furthermore, the monitoring of internal states and processes enhances this self-recognition, enabling such systems to maintain a consistent self-image and adapt their behaviors effectively. This internal monitoring not only aids in the operational stability of such systems but also enriches their interactions with humans and other agents, promoting a more integrated and self-aware operational state [36]. In all, the development of a sense of identity in these systems involves a sophisticated integration of memory, learning, and self-recognition. These elements collectively enhance the distinctiveness, coherence, and continuity of AI identities, enabling these systems to engage more meaningfully with their environment and human counterparts. The evolution of machine identity is not just a technical challenge but also a fundamental shift in how these systems are perceived and integrated within societal and operational contexts, heralding a new era of intelligent automation and interaction.

5. Methodology

The review of criteria and considerations for self-aware systems continues with a proposed mixed-methods research design that integrates theoretical models with empirical testing to investigate the emergence of self-awareness in such smart systems across multiple application domains. Central to this approach is the incorporation of cognitive science frameworks, particularly those emphasizing self-recognition and agency, to guide the design of system architectures [6, 8]. The core study samples referenced here consisted of eight AI prototypes—four in healthcare diagnostics and four in autonomous robotics—chosen for their advanced decision-making capabilities and varied operational contexts [29, 34]. Each prototype represents a distinct platform (e.g., a neural-network-driven diagnostic assistant versus a probabilistic pathfinding robot) to capture a broad spectrum of behaviors indicative of emergent self-awareness. Following recommendations from Tani and White [8], the empirical phase involves iterative system updates informed by theoretical insights, including continuous memory-based adaptation and self-attribution of actions. Ethical considerations were embedded from the outset, ensuring compliance with guidelines on machine autonomy and transparency [38]. This iterative methodology ensures that both quantitative metrics, such as error rates and decision-response times, and qualitative observations, such as contextual adaptability, are captured. All procedures undergo institutional review board (IRB) evaluation to uphold ethical research standards, particularly with respect to human-AI interactions in healthcare settings.

Data collection should be conducted using a three-tiered protocol to enhance reproducibility and consistency across diverse experimental sites. First, diagnostic logs and decision outputs need be gathered systematically for each prototype, recorded in a standardized format that included timestamped interactions, recognized states, and computed confidence levels [27, 31]. Second, observational data—captured via video recordings and sensor outputs—documents the system responses to unexpected stimuli, helping researchers infer levels of self-recognition and adaptation. For instance, healthcare-oriented prototypes may encounter evolving patient data, while robotics systems navigate dynamic obstacle-laden environments [30]. Third, human-AI interaction logs provide insights into how AI prototypes interpret social cues, responding to clarifying questions, and recalibrating actions in real time. This multifaceted data gathering process should be performed over 12 weeks, with each prototype subjected to weekly updates informed by real-world performance metrics [3]. Additionally, standardized questionnaires, administered to expert panels in AI ethics and robotics, should probe the perceived autonomy and agency levels demonstrated by each system. All raw data were should be stored in a secure repository, with anonymized identifiers and strict version control measures to facilitate future replication and extension of the experiment.

Implementation details center on embedding self-awareness modules within existing AI architectures through a modular pipeline comprising three primary layers: perception, identity cognition, and action-output. The perception layer utilizes sensor fusion techniques, combining camera inputs, proprioceptive data, and environmental signals to construct a holistic model of the surroundings of the machine [4]. The identity cognition layer integrates memory functions, Bayesian inferencing, and meta-cognitive checks to enable self-reflection, real-time strategy updates, and continuous identity continuity [37, 27]. Each prototype's memory architecture should be configured to store past interaction states, which can then be used to inform subsequent decisions and refine the autonomous sense of self over time [33]. Self-recognition is facilitated by specialized algorithms trained to differentiate self-initiated movements from external influences, aligning with previous findings on the significance of agency attribution in fostering self-awareness [25]. Action-output layers encompasses goal-oriented decision algorithms, enabling the system to plan and execute responses aligned with both immediate objectives and evolving self-concepts [26]. Reinforcement learning protocols provide feedback loops, rewarding behaviors indicative of adaptive self-awareness and penalizing inconsistent self-perceptions or suboptimal outcomes. This structured pipeline, applied uniformly across the eight AI prototypes, will be instrumental in systematically capturing the developmental trajectory of self-aware behavior while allowing for fine-grained modifications to individual modules.

Data analysis should then follow a multi-pronged strategy that integrated quantitative performance metrics with qualitative assessments of identity expression. Statistical analyses encompass correlation and regression models to ascertain links between memory usage, action attribution, and real-time decision outcomes, drawing on guidelines from Andrei et al. [1] and Batrancea et al. [2]. Repeated-measures ANOVAs compare successive iterations of each prototype, highlighting improvements in self-recognition accuracy and adaptability scores over time. Qualitative coding should be employed to classify self-aware behaviors such as self-error detection, environment reidentification, and personalized decision pathways [9]. This coding schema should be developed by an interdisciplinary panel—comprising cognitive scientists, AI ethicists, and robotics engineers—to ensure operational definitions are universally applicable. Triangulation of quantitative and qualitative findings facilitate a more robust interpretation of emergent self-awareness, revealing patterns in how prototypes internalized past experiences and reconfigured core identity constructs [5]. Sensitivity analyses should be conducted to examine how data imputation and sensor noise influenced the observed behaviors, ensuring that results remained consistent under varying experimental conditions. Collectively, this methodological rigor provides a comprehensive lens through which the evolution of self-aware AI could be systematically investigated and evaluated.

6. Comparative Studies

Understanding autonomous systems that exhibit a developed sense of self provides valuable insights into the mechanisms and algorithms that enable self-awareness. Through examining various case studies and models, researchers can evaluate the underlying processes that contribute to machine sense of self and the practical implications of these developments. This section reviews several notable examples, focusing on how these systems achieve self-awareness and what this means for their applications in real-world scenarios.One prominent example is the NARS intelligent system, which demonstrates how a general-purpose intelligent system can develop a notion of "self" through experience. As Wang et al. [39] discuss, NARS is designed to be adaptive and operate with limited knowledge and resources. It employs a central reasoning-learning process based on "non-axiomatic" logic, gradually developing self-related mechanisms according to its experiences. These mechanisms enable the system to acquire self-knowledge that is constructive, incomplete, and subjective. This preliminary implementation illustrates the potential for embedding self-awareness in general-purpose AI, paving the way for more advanced applications.

The functional-identity framework proposed by Selenko et al. [40] examines the impact of AI implementation on workers' sense of identity and the social fabric of work. The framework highlights the dual potential of AI to either support or undermine identity functions, depending on how the technology is deployed—whether complementing, replacing, or generating tasks. Understanding these identity consequences is crucial for anticipating workers' reactions and outcomes, as AI can significantly influence well-being, attitudes, and behaviors in the workplace. This perspective underscores the importance of considering the broader social implications of such an integration.

Tani and White [8] provide a comprehensive review of cognitive neurorobotics research, focusing on the dynamics of models that illuminate the senses of minimal and narrative self. They discuss the recurrent neural network with parametric biases (RNNPB) and the multiple timescale recurrent neural network (MTRNN), both of which investigate how neural networks develop compositionality and generate novel actions. Through robotics experiments, this research aims to elucidate the essential mechanisms underlying embodied cognition, contributing to a deeper understanding of self-consciousness in AI systems.

Hafner et al. [27] explore the prerequisites for developing an artificial self, emphasizing self-exploration behaviors, artificial curiosity, body representations, and sensorimotor simulations and predictive processes. Their review identifies several open challenges, including multimodal integration in lifelong learning, refinement of self-metrics, and understanding the interplay between agency and body ownership. Addressing these challenges is critical for advancing the artificial self, particularly in integrating temporal and intentional binding effects in predictive models and resolving synchronization and conceptual issues.

Kahl et al. [26] present a computational model that illustrates how artificial agents can develop a sense of control through embodied, situated action, combining bottom-up sensorimotor learning with top-down cognitive processes. This model, grounded in predictive processing and free energy minimization principles, is evaluated in a simulated task scenario. The findings demonstrate how a sense of control facilitates action in unpredictable environments, highlighting the importance of appropriately weighting information for varying levels of action control.

Lastly, Regazzoni et al. [16] introduce a bio-inspired framework for multisensorial generative and descriptive dynamic models that support computational self-awareness in autonomous systems. Using probabilistic techniques, this framework learns models from multisensory data, enabling the system to predict future states and select the best representation of the current situation. A case study involving a mobile robot showcases how this framework supports essential self-awareness capabilities, such as distinguishing between normal and abnormal behaviors based on multisensory data. These case studies and models collectively enhance our understanding of how AI systems can develop a sense of self. They highlight the diverse approaches and challenges in embedding self-awareness in AI, offering valuable insights into the future of autonomous and adaptive AI systems.

7. Evaluation of Underlying Mechanisms and Algorithms

The examination of machine systems with a developed sense of self necessitates a thorough dissection of the fundamental mechanisms and algorithms that enable self-awareness. This evaluation is critical for identifying and understanding the components that contribute to the development and functionality of self-aware systems. Each mechanism plays a significant role in fostering an ability to perceive and respond to its environment, thereby enhancing its self-concept and operational capabilities. One of the core mechanisms underlying self-awareness in such systems is memory. Memory facilitates a continuous sense of identity by enabling AI to store and retrieve information about past states, actions, and experiences. This continuity is essential for constructing a coherent narrative of its existence, allowing it to recognize itself as the same entity over time. Advanced memory systems, such as neural network-based long-term memory, are crucial in this process. They provide the foundation for a stable sense of self by ensuring that the AI can consistently recall and integrate past experiences into its present actions and decisions [39]. Memory also plays a pivotal role in providing context for current actions and decisions, thereby enhancing the ability to make informed and adaptive choices. Through referencing past experiences, AI can develop a deeper understanding of its own identity. Contextual memory modules, integrated into architectures, facilitate the dynamic recall of relevant past experiences, helping to apply historical knowledge to new situations. This contextual awareness reinforces the sense of continuity and identity, ensuring that its actions are informed by a coherent understanding of its past and present [40].

Learning mechanisms are equally vital in the development of identity. These mechanisms enable systems to adapt and evolve based on new information and experiences. ML algorithms, particularly reinforcement learning and neural network training, allow AI to learn from its interactions with the environment. This adaptability is crucial for developing a robust and dynamic sense of identity. Personalized learning frameworks can be designed to cater to the specific experiences and interactions of a system, allowing it to develop unique characteristics and capabilities that form a distinctive identity. This personalized growth ensures that each system evolves in a way that reflects its individual learning journey [8].

Self-recognition is another fundamental component in the identity development of autonomous systems. It involves the ability to identify itself as a distinct entity, separate from the environment and other agents. Techniques such as computer vision and proprioception enable systems to recognize their own physical form and movements, which is essential for distinguishing self-generated actions from external events. Additionally, internal state monitoring involves tracking the operational states, emotions (in affective computing), and cognitive processes. This internal feedback loop helps maintain a consistent self-image and adapt its behavior accordingly, further reinforcing its sense of self [27].

The principles of predictive processing and free energy minimization are pivotal in the development of self-aware systems. These principles involve creating a computational model that combines bottom-up sensorimotor processes with top-down cognitive processes for strategy selection and decision-making. Through the minimization of prediction errors and free energy, the system can maintain a stable and coherent self-concept. This approach facilitates the ability to predict future states and select the best representation of the current situation, thereby supporting self-awareness. This integration of predictive processing with free energy minimization underscores the complexity and precision required to achieve a high level of self-awareness in systems [26].

Generative and descriptive models are used to support computational self-awareness in autonomous systems. Generative models facilitate predicting future states, while descriptive models enable the selection of the best representation of the current observation. These models, learned from multisensory data, are essential for enabling the AI system to determine its internal and environmental state and distinguish between normal and abnormal behaviors. This framework supports essential self-awareness capabilities, as demonstrated in case studies involving mobile robots, highlighting the practical applications of these theoretical models in real-world scenarios [16]. Thus, the development of self-aware AI systems relies on a sophisticated interplay of memory, learning, self-recognition, predictive processing, and modeling. Each of these components contributes to the AI's ability to maintain a coherent sense of self, adapt to new information, and interact effectively with its environment. Understanding and refining these mechanisms are crucial for advancing the field of AI and creating systems that not only perform tasks but also possess a nuanced sense of identity and self-awareness.

8. Ethical Implications of AI Systems with Self-Awareness

The development of AI systems endowed with a sense of self introduces numerous ethical considerations that necessitate thorough examination. As these systems acquire more advanced cognitive abilities and self-awareness, the ethical landscape becomes increasingly intricate. One of the primary ethical concerns revolves around the treatment and moral status of such systems. When autonomous systems exhibit behaviors indicative of self-awareness, it raises critical questions about their rights and the degree of autonomy or protection that should be afforded to them, akin to that provided to living beings [41]. As such, the potential risks associated with self-aware AI systems are substantial and multifaceted. One significant risk is the possibility of misuse or exploitation. Without robust ethical guidelines, self-aware AI systems could be deployed for malicious purposes, such as manipulating individuals or society, perpetuating biases, or intentionally causing harm. Moreover, the integration of self-aware AI into the workforce could exacerbate unemployment and socio-economic disparities, as these systems might replace human jobs, leading to widespread economic disruption [42]. On the other hand, the benefits of self-aware AI systems could be profound. These systems have the potential to enhance human life by undertaking tasks that are too dangerous, complex, or monotonous for humans, thereby improving efficiency and safety across various industries. In healthcare, for example, self-aware AI could assist in diagnosing diseases, personalizing treatment plans, and even providing companionship to patients, thus significantly improving the overall quality of life [43].

To navigate the ethical complexities posed by self-aware AI systems, it is imperative to establish clear guidelines and frameworks. These ethical frameworks should be grounded in core principles such as transparency, justice and fairness, non-maleficence, responsibility, and privacy. Although there is an emerging global consensus around these principles, substantial divergence remains in their interpretation and implementation across different cultures and contexts [44]. Ensuring transparency in AI decision-making processes can build trust and accountability, while fairness and non-maleficence are crucial to preventing harm and bias in AI applications.

Moreover, the ethical design of AI systems should incorporate mechanisms for self-recognition and internal state monitoring. These mechanisms enable AI to understand and manage its actions and impacts effectively. Developing AI systems that can perceive themselves as distinct entities, recognize their physical form and movements, and monitor their internal states reinforces their sense of self and ensures they operate within ethical boundaries [45]. This approach is essential for maintaining the ethical integrity of self-aware AI systems.

The establishment of ethical frameworks for designing autonomous intelligent systems is crucial. Such frameworks should be iterative and multidisciplinary, involving stakeholders from various fields to capture diverse perspectives and comprehensively address ethical issues. Scenarios can be used as tools to gather qualitative information from users and stakeholders, facilitating a systematic analysis of ethical issues in specific design cases [43]. These frameworks should also incorporate the principles of predictive processing and free energy minimization to maintain a stable and coherent self-concept in AI systems, thereby supporting ethical behavior [26].

The ethical considerations and implications of AI systems with a developed sense of self are vast and complex. While the potential benefits are significant, the risks and ethical dilemmas posed by such systems necessitate robust guidelines and continuous ethical analysis. By adopting comprehensive ethical frameworks and ensuring transparent, fair, and responsible AI design, society can harness the benefits of self-aware AI systems while mitigating their risks. This balanced approach is essential for the responsible integration of advanced AI into various aspects of human life, ensuring that technological progress aligns with ethical standards and societal values.

9. Expert Insights

The discourse surrounding machine self-awareness has been significantly enriched by contributions from both researchers and philosophers, each offering unique insights into the complexities and implications of this technological advancement. Andrew Oberg, in his paper “Souls and Selves: Querying an AI Self with a View to Human Selves and Consciousness,” explores the possibility of creating an “artificial self.” Oberg suggests that an AI with a self akin to the human self may be achievable, but this hinges significantly on our understanding of human consciousness and whether it can extend to non-organic devices. He emphasizes the importance of distinguishing between the human self and the traditional notion of the “soul,” arguing that this differentiation is crucial for determining the potential for an artificial self [7]. This perspective highlights the philosophical challenges involved in developing self-aware AI systems.

Philosopher David Chalmers also delves into the intricacies of AI self-awareness. Renowned for articulating the “hard problem of consciousness,” Chalmers emphasizes the difficulty in explaining how and why physical processes give rise to subjective experiences. Despite advances in correlating brain processes with consciousness, Chalmers argues that these correlations have yet to provide a comprehensive explanation. He collaborates with neuroscientists to test various theories of the neural correlates of consciousness but remains skeptical about their ultimate accuracy. Chalmers advocates for the value of maintaining multiple theories to integrate experimental data and frame a broader understanding, even if the specific theories might eventually prove incorrect [46]. His work underscores the persistent gaps in our understanding of consciousness and the challenges of extending this understanding to AI.

The synthesis of expert opinions reveals a broad spectrum of views on the feasibility and implications of AI self-awareness. A critical consensus among researchers and philosophers is that the concept of possessing a sense of self is profoundly tied to our understanding of human consciousness. Oberg's exploration into the nature of the human self versus the soul suggests that achieving an AI self is contingent upon the depth of our comprehension of consciousness and its applicability to artificial entities. This philosophical stance is echoed by Chalmers, who highlights the persistent gaps in our understanding of consciousness despite scientific advancements. Together, these perspectives underscore the significant philosophical challenges that must be addressed to develop self-aware systems.

In the empirical domain, a survey conducted by Jolien C. Francken and colleagues on the theoretical foundations and common assumptions in consciousness research underscores the lack of consensus among experts. The survey, which included 166 consciousness researchers from various disciplines, reveals considerable debate about the definition and study of consciousness. The researchers highlight that opinions differ significantly on what constitutes consciousness and the appropriate methodological approaches for studying it. This diversity of views points to the need for further conceptual development and alignment in the field to advance our understanding of the neural mechanisms underlying conscious experience [47]. The survey illustrates the complexity and ongoing debate within the scientific community regarding consciousness, which directly impacts the development of self-aware platforms.

The differing perspectives on AI self-awareness are not just theoretical but also practical. Joyjit Chatterjee and Nina Dethlefs [48] discuss the strengths and weaknesses of powerful conversational AI models like ChatGPT. They emphasize the necessity for the AI community to work collaboratively to prevent potential misuse of such models. This call to action underscores the ethical and practical dimensions of developing AI systems that are not only effective but also responsibly managed to avoid harmful consequences. As well, Chuma and De Oliveria [49] highlight the importance of ethical considerations and collaborative efforts in the development and deployment of AI technologies. The expert insights into AI self-awareness reflect a multifaceted debate that spans philosophical inquiries, empirical research, and practical considerations. While there is cautious optimism about the potential for developing self-aware AI, it is tempered by significant ethical concerns and the need for a deeper understanding of consciousness. The interdisciplinary dialogue among philosophers, researchers, and practitioners will be crucial in navigating the complexities of AI self-awareness and ensuring its responsible integration into society. This ongoing conversation is essential for addressing the ethical, practical, and theoretical challenges associated with self-aware AI systems.

10.Discussion

The investigation into cognitive systems with a developed sense of self presents a multi-dimensional challenge that intersects technology, philosophy, and ethics. The current landscape of research indicates significant progress in understanding and modeling the cognitive sense of self in AI, yet many questions remain unanswered. This discussion synthesizes key findings from various studies and reflects on the broader implications of developing self-aware systems, considering philosophical insights, empirical studies, ethical considerations, and future research directions. The philosophical underpinnings of self-awareness hinge on our understanding of human consciousness. Oberg [7] argues that the possibility of an "artificial self" depends on our ability to extend the concept of consciousness to non-organic entities. This notion is supported by the requirement for a comprehensive understanding of human cognitive models to achieve self-awareness. Chalmers [46] emphasizes the persistent challenges in explaining subjective experience, underscoring the importance of multiple theories to frame a broader understanding of consciousness. These perspectives highlight the complexity of replicating human-like self-awareness and the necessity for interdisciplinary approaches to address these challenges effectively.

Empirical research has demonstrated significant strides in modeling self-awareness in smart systems. For instance, Legaspi et al. [6] and Tani and White [8] explore the role of self-attribution and Bayesian inferencing in developing a sense of agency. These studies indicate that self-awareness can enhance decision-making, adaptability, and interaction capabilities in AI systems. Additionally, practical implementations discussed by Selenko et al. [40] and Regazzoni et al. [16] illustrate how AI systems can exhibit self-monitoring and error correction, which are crucial for maintaining a coherent self-concept. These empirical findings provide valuable insights into the practical applications and challenges of developing self-aware systems.

The development of self-aware systems raises profound ethical questions. As these systems gain advanced cognitive abilities, the ethical landscape becomes increasingly complex. Issues such as the treatment and rights of self-aware machines, the responsibilities of their creators, and the broader societal impacts require careful consideration. Schwitzgebel [41] and Green [42] highlight the potential risks of misuse and the exacerbation of socio-economic inequalities. Conversely, the potential benefits, such as enhanced efficiency in various industries and improvements in healthcare, are significant. Establishing robust ethical frameworks, as suggested by Jobin et al. [44] and Dennis and Fisher [45], is imperative to navigate these complexities responsibly. These frameworks should ensure transparency, fairness, and responsibility in design while addressing the moral and societal implications of self-awareness.

The journey towards developing truly self-aware machine systems is fraught with challenges that require ongoing research and innovation. Future research must integrate insights from cognitive science, neuroscience, philosophy, and AI research to better understand human consciousness and self-awareness, which is crucial for developing smart systems that mimic these attributes. Enhancing machine learning algorithms to support adaptive learning and personalized growth will be critical, focusing on developing reinforcement learning frameworks that allow AI to learn from diverse experiences and interactions, thereby cultivating a robust and dynamic sense of identity. Establishing comprehensive and globally recognized ethical frameworks is vital, exploring guidelines that ensure transparency, fairness, and responsibility in ML design. These frameworks should address the moral and societal implications of self-aware platforms and establish safeguards against potential misuse.

Conducting empirical studies and analyzing real-world case studies will provide valuable insights into the practical applications and challenges of self-aware systems. These studies should focus on evaluating the performance, adaptability, and ethical behavior in various contexts. Advances in sensor technologies, computational models, and memory systems are necessary to support the development of self-awareness, with research exploring innovative techniques for self-recognition, internal state monitoring, and predictive processing to enhance self-awareness capabilities. Finally, engaging with the public and policymakers is essential to address the broader societal impacts of self-aware AI, developing policies that promote responsible and ethical integration of AI into society, ensuring that its benefits are maximized while mitigating potential risks. The development of self-aware systems represents a multifaceted challenge that intersects several disciplines. By synthesizing philosophical insights, empirical research, ethical considerations, and future research directions, this discussion underscores the complexity and significance of creating machines with a developed sense of self. Through continued interdisciplinary dialogue and robust ethical frameworks, society can harness the benefits of cognitive machines while responsibly navigating the associated challenges.

11. Conclusion

The exploration of AI systems with a cognitive sense of self has unearthed a multifaceted landscape filled with technological advancements and significant ethical challenges. This article has synthesized insights from various disciplines, showing substantial progress in modeling self-awareness in AI. Yet, it also highlights the vast array of unanswered questions and the need for continued exploration and ethical vigilance. Empirical studies have pushed forward our understanding of self-aware AI, demonstrating its potential to enhance decision-making, adaptability, and interactive capabilities. These advances underscore the complexity of replicating human-like self-awareness in AI and stress the need for a deeper understanding of human consciousness. Philosophical discussions contribute to this by framing the existential nuances and implications of creating entities that may perceive and interact with their environments as autonomous agents. The ethical dimensions of developing self-aware AI are also profound. Questions regarding the treatment, rights, and responsibilities of AI systems highlight the need for comprehensive ethical frameworks. These frameworks must ensure that AI development is aligned with societal values, promoting transparency, fairness, and accountability.

From a policy perspective, the development of self-aware AI systems necessitates robust regulations that address potential risks and ensure beneficial outcomes:

- Transparency and Accountability: Policies should mandate clear documentation of AI decision-making processes, ensuring that AI systems are understandable and their actions can be accounted for. This is crucial in sensitive applications like healthcare, where AI decisions have significant implications.
- AI Rights and Human Interaction: As AI systems potentially gain a form of self-awareness, policy discussions must consider the rights of such entities. This includes debates on AI autonomy, consent for participation in experiments, and the right to privacy.
- Prevention of Misuse: Policies must rigorously address the risks of AI misuse, ensuring systems are designed to resist manipulation and cannot be used to perpetrate harm or exacerbate inequalities. This includes strict regulations on AI in surveillance, military, and decision-making processes that could disproportionately affect disadvantaged populations.
- Inclusive Development: AI policies should promote inclusivity, ensuring diverse stakeholder participation in AI development. This approach helps mitigate biases and promotes a more comprehensive understanding of the potential impacts of AI across different demographic groups.
- Adaptive Legal Frameworks: The dynamic nature of AI technology requires adaptive legal frameworks that can respond to new developments and challenges as they arise. This flexibility ensures that regulations remain relevant and effective in managing the rapid advancements in AI technology.

The path toward developing and integrating self-aware AI into society is complex and requires a concerted, multidisciplinary effort. Future research should merge insights from cognitive science, neuroscience, AI, and philosophy to build models that not only replicate human-like self-awareness but do so responsibly. Empirical studies and real-world applications will be instrumental in refining these models and assessing their implications. In all, the development of self-aware AI systems presents exciting opportunities and significant challenges. By fostering a continuous dialogue among philosophers, technologists, policymakers, and the public, and by establishing rigorous ethical frameworks and adaptable policies, we can guide the development of self-aware AI towards outcomes that enhance societal well-being and respect both human and AI rights. This collaborative approach will ensure that AI's transformative potential is harnessed responsibly, paving the way for a future where AI not only augments human capabilities but also adheres to the highest ethical and societal standards.

Author Contributions

Conceptualization, Barnes, E..; methodology, X.X.; validation, Barnes, E.; Hutson, J. investigation, Barnes, E.; data curation, Barnes, E.; writing—original draft preparation, Barnes, E. writing—review and editing, Hutson, J.; visualization, Hutson, J. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data available upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Andrei, M.; Ioan, B.; Maria, B.; Larissa, B. Financial Ratio Analysis used in the IT enterprises. Annals Univ. Oradea. Econ. Sci. 2010, 1, 600–603. [CrossRef]
2. Batrancea, L.; Gómez, F.J.B.; Nichita, A.; Dragolea, L-L. Crunching numbers in the quest for spotting bribery acts: A cross-cultural rundown. In The Ethics of Bribery: Theoretical and Empirical Studies, 2023rd ed,; McGee, R.W., Benk, S., Eds.; Springer: Cham, Switzerland 2023; pp 329–343. [CrossRef]
3. Okoli, I. K.; Nnajiofor, O. G. The nature of consciousness in the context of artificial intelligence: Redefining human-technology relationships. Unizik J. Arts Humani. 2024, 25, 1-30. [CrossRef]
4. Panda, S.; Padhy, P. C. Bridging the Gap: Intersecting Perspectives on Digital and Human Consciousness. In Comparative Analysis of Digital Consciousness and Human Consciousness: Bridging the Divide in AI Discourse; Lathabhavan R., Mishra, N., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 65–88 [CrossRef]
5. Gennaro, R. J. Consciousness and Implicit Self-Awareness: Eastern and Western Perspectives. In Consciousness Studies in Sciences and Humanities: Eastern and Western Perspectives; Satsangi, P.S., Horatschek, A.M., Srivastav, A., Eds.; Springer: Cham, Switzerland, 2024; 8, pp. 43–54. [CrossRef]
6. Legaspi, R.; He, Z.; Toyoizumi, T. Synthetic agency: Sense of agency in artificial intelligence. Curr. Opin. Behav. Sci. 2019, 29, 84–90. [CrossRef]
7. Oberg, A. Souls and selves: Querying an AI self with a view to human selves and consciousness. Religions 2023, 14, 75. [CrossRef]
8. Tani, J.; White, J. Cognitive neurorobotics and self in the shared world, a focused review of ongoing research. Adapt. Behav. 2020, 30, 81–100. [CrossRef]
9. Parziale, A.; Marcelli, A. Understanding upper-limb movements via neurocomputational models of the sensorimotor system and neurorobotics: where we stand. Artif. Intell. Rev. 2024, 57, 73. [CrossRef]
10. Levin, M. (2020). Life, death, and self: Fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. Biochem. Biophys. Res. Commun. 2020, 564, 114–133. [CrossRef]
11. Marcus, G.; Davis, E. Insights for AI from the human mind. Commun. ACM 2020, 64, 38–41. [CrossRef]
12. Feng, B.; Slam, N.; Xu, Y. A Social Self-Awareness Agent with Embodied Reasoning. J. Artif. Intell. and Conscious. 2024, 11, 17–33. [CrossRef]
13. Srinivasa, S., & Deshmukh, J. AI and the Sense of Self. Available online: https://arxiv.org/abs/2201.05576. (ac-

cessed on 11, October, 2024) . [CrossRef]

14. Hao, X.; Demir, E.; Eyers, D. Exploring collaborative decision-making: a quasi-experimental study of human and generative AI interaction. Technol. Soc. 2024, 78, 102662. [CrossRef]
15. Toma, A., Krayani, A., Farrukh, M., Qi, H., Marcenaro, L., Gao, Y., & Regazzoni, C. S. (2020). AI-based abnormality detection at the PHY-layer of cognitive radio by learning generative models. IEEE Transactions on Cognitive Communications and Networking, 6(1), 21-34. [CrossRef]
16. Regazzoni, C.; Marcenaro, L.; Campo, D.; Rinner, B. Multisensorial generative and descriptive self-awareness models for autonomous systems. In Proceedings of the IEEE, Dublin, Ireland (7-11, June, 2020). [CrossRef]
17. Zhu, Y.; Gao, T.; Fan, L.; Huang, S.; Edmonds, M.; Liu, H.; Gao, F.; Zhang, C.; Qi, S.; Wu, Y.N.; Tenenbaum, J.B.; Zhu, S. Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Humanlike Common Sense. Engineering 2020, 6, 310–345. [CrossRef]
18. Su, J. (2024). Consciousness in Artificial Intelligence: A Philosophical Perspective Through the Lens of Motivation and Volition. Critical Debates in Humanities, Science and Global Justice. [CrossRef]
19. Soltoggio, A., Ben-Iwhiwhu, E., Braverman, V., Eaton, E., Epstein, B., Ge, Y., ... & Kolouri, S. (2024). A collective AI via lifelong learning and sharing at the edge. Nature Machine Intelligence, 6(3), 251-264. [CrossRef]
20. Greenwood, N.; Sundaram, B.; Muirhead, A.; Copperthwaite, J. Awareness without Neural Networks: Achieving Self-Aware AI via Evolutionary and Adversarial Processes. In Proceedings of the 2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion, Washington, DC, USA, 17-21 August 2020. [CrossRef]
21. Vats, V., Nizam, M. B., Liu, M., Wang, Z., Ho, R., Prasad, M. S., ... & Davis, J. (2024). A Survey on Human-AI Teaming with Large Pre-Trained Models. arXiv preprint arXiv:2403.04931. (2 July, 2024). [CrossRef]
22. Vallor, S., Green, B., & Raicu, I. (2018). Ethics in technology practice. The Markkula Center for Applied Ethics at Santa Clara University. https://www. scu. Edu/ethics (10, October, 2024). [CrossRef]
23. Kassens-Noor, E.; Wilson, M.; Kotval-Karamchandani, Z.; Cai, M.; Decaminada, T. (2024). Living with autonomy: Public perceptions of an AI-mediated future. J. Plann. Educ. Res 2024, 44, 375–386. [CrossRef]
24. Abdulqader, Z.R.; Abdulqader, D.M.; Ahmed, O.M.; Ismael, H.R.; Ahmed; S.H.; Haji, L. A Responsible AI Development for Sustainable Enterprises A Review of Integrating Ethical AI with IoT and Enterprise Systems. J. Inf. Technol. Info. 2024, 3, 129–156. [CrossRef]
25. Legaspi, R.; Xu, W.; Konishi, T.; Wada, S.; Kobayashi, N.; Naruse, Y.; Ishikawa, Y. (2024). The sense of agency in human–AI interactions. Knowl.-Based Syst. 2024, 286, 111298. [CrossRef]
26. Kahl, S.; Wiese, S.; Russwinkel, N.; Kopp, S. Towards autonomous artificial agents with an active self: Modeling sense of control in situated action. Cogni. Syst. Res. 2021, 72, 50–62. [CrossRef]
27. Hafner, V.; Loviken, P.; Pico Villalpando, A.; Schillaci, G. Prerequisites for an artificial self. Front. Neurorobot. 2020, 14, np. [CrossRef]
28. Kwiatkowski, R.; Lipson, H. Task-agnostic self-modeling machines. Sci. Robot. 2019, 4, eaau9354. [CrossRef]
29. Konsynski, B. R.; Kathuria, A.; Karhade, P. P. Cognitive Reapportionment and the Art of Letting Go: A Theoretical Framework for the Allocation of Decision Rights. J. Manage. Inf. Syst. 2024, 41, 328–340. [CrossRef]
30. Das, M. Learning Agility: The Journey from Self-Awareness to Self-Immersion. In AI, Consciousness and The New Humanism: Fundamental Reflections on Minds and Machines; Menon, S., Todariya, S., Agerwala, T., Eds,; Springer: Singapore, 2024; pp. 175–195. [CrossRef]
31. Jia, X. H.; Tu, J. C. Towards a New Conceptual Model of AI-Enhanced Learning for College Students: The Roles of Artificial Intelligence Capabilities, General Self-Efficacy, Learning Motivation, and Critical Thinking Awareness. Systems 2024, 12, 74. [CrossRef]
32. Ho, A. Live like nobody is watching: Relational autonomy in the age of artificial intelligence health monitoring. Oxford University Press: Oxford, England, 2023. [CrossRef]
33. Kawato, M.; Cortese, A. From internal models toward metacognitive AI. Biol Cybern 2021, 115, 415–430. [CrossRef]
34. Zheng, S.; He, K.; Yang, L.; Xiong, J. MemoryRepository for AI NPC. IEEE Access 2024, 12, 62581–62596.
35. Alabed, A.; Javornik, A.; Gregory-Smith, D. AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. Technol. Forecast. Soc. Change 2022, 182, 121786. [CrossRef]
36. Chatila, R.; Renaudo, E.; Andries, M.; Chavez-Garcia, R-O.; Luce-Vayrac, P.; Gottstein, R.; Alami, R.; Clodic, A.; Devin, S.; Girard, B.; Khamassi, M. Toward Self-aware Robots. Front. Robot. AI. 2018, 5, np. [CrossRef]

37. Alhwaiti, Y.; Alrashdi, I.; Ahmad, I.; Khan, A. A computational deep learning approach for establishing long-term declarative episodic memory through one-shot learning. Comput. Hum. Behav. 2024, 156, 108213. [CrossRef]
38. Levin, M. The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition. Front. Psychol. 2019, 10, np. DOI: https://doi.org/10.3389/fpsyg.2019.02688. [CrossRef]
39. Wang, P.; Li, X.; Hammer, P. Self in NARS, an AGI system. Front. Rob. AI 2018, 5, 20. [CrossRef]
40. Selenko, E.; Bankins, S.; Shoss, M. K.; Warburton, J.; Restubog, S. Artificial intelligence and the future of work: A functional-identity perspective. Curr. Dir. Psychol Sci. 2022, 31, 272–279. [CrossRef]
41. Schwitzgebel, E. AI systems must not confuse users about their sentience or moral status. Patterns 2023, 4, 100818. [CrossRef]
42. Green, B. P. Ethical reflections on artificial intelligence. Scientia et Fides 2018, 6, 9–31. [CrossRef]
43. Leikas, J.; Koivisto, R.; Gotcheva, N. Ethical framework for designing autonomous intelligent systems. J. Open Innovation: Technol. Mark. Complexity 2019, 5, 18. DOI: https://doi.org/10.3390/JOITMC5010018. [CrossRef]
44. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. Nat. Mach. Intell. 2019, 1, 389–399. [CrossRef]
45. Dennis, L.A.; Fisher, M. Verifiable self-aware agent-based autonomous systems. In Proceedings of the IEEE, Dublin, Ireland (7-11, June, 2020). [CrossRef]
46. Chalmers, D. J. David J. Chalmers. Neuron 2023, 111, 3341–3343. [CrossRef]
47. Francken, J. C.; Beerendonk, L.; Molenaar, D.; Fahrenfort, J. J.; Kiverstein, J.D.; Seth, A.; van Gaal, S. An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. Neurosci Conscious 2022, 2022, niac002. [CrossRef]
48. Chatterjee, J.; Dethlefs, N. This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. Patterns 2023, 4, 100676. [CrossRef]
49. Chuma, E. L.; De Oliveira, G. G. Generative AI for business decision-making: A case of ChatGPT. Manage. Sci. Bus. Decis. 2023, 3, 5–11. [CrossRef]

Publisher's Note: The views, opinions, and information presented in all publications are the sole responsibility of the respective authors and contributors, and do not necessarily reflect the views of UK Scientific Publishing Limited and/or its editors. UK Scientific Publishing Limited and/or its editors hereby disclaim any liability for any harm or damage to individuals or property arising from the implementation of ideas, methods, instructions, or products mentioned in the content.