Beyond Human and Machine: An Architecture and Methodology Guideline for Centaurian Design

Remo Pareschi

STAKE Lab, University of Molise, 86100 Campobasso, Italy; remo.pareschi@unimol.it

Abstract: The concept of the centaur, symbolizing the fusion of human and machine intelligence, has intrigued visionaries for decades. Recent advancements in artificial intelligence have made this concept not only realizable but also actionable. This synergistic partnership between natural and artificial intelligence promises superior outcomes by leveraging the strengths of both entities. Tracing its origins back to early pioneers of human–computer interaction in the 1960s, such as J.C.R. Licklider and Douglas Engelbart, the idea initially manifested in centaur chess but faced challenges as technological advances began to overshadow human contributions. However, the resurgence of generative AI in the late 2010s, exemplified by conversational agents and text-to-image chatbots, has rekindled interest in the profound potential of human–AI collaboration. This article formalizes the centaurian model, detailing properties associated with various centaurian designs, evaluating their feasibility, and proposing a design methodology that integrates human creativity with artificial intelligence. Additionally, it compares this model with other integrative theories, such as the Theory of Extended Mind and Intellectology, providing a comprehensive analysis of its place in the landscape of human–machine interaction.

Keywords: centaurian models; hybrid intelligence; man–machine interaction; design science

1. Introduction

In the evolving landscape of technology and human interaction, hybrid intelligence systems integrating human cognitive abilities with artificial intelligence (AI) are gaining significant interest. These systems aim to enhance overall performance and capabilities by leveraging the strengths of both human and artificial components. This article explores centaurian models of hybrid intelligence, which uniquely blend these components to exploit their complementary strengths.

The integration of human and machine intelligence has deep historical roots, articulated by pioneers such as Douglas Engelbart and J.C.R. Licklider. Engelbart, known for inventing the computer mouse and his work in interactive computing, presented *The Mother of All Demos* in 1968, showcasing real-time text editing, video conferencing, and hypertext. In his 1962 article *Augmenting Human Intellect: A Conceptual Framework* [1], he defined augmentation as enhancing a person's ability to solve complex problems and argued that computers could amplify human cognitive abilities by displaying and manipulating information and providing interactive feedback.

Before Engelbart, J.C.R. Licklider, in his 1960 article *Man-Computer Symbiosis* [2], defined symbiosis as the mutually beneficial coexistence of different organisms, applying this concept to human-computer relationships. He envisioned that combining the strengths and compensating for the weaknesses of humans and computers would result in higher performance and intelligence than either could achieve alone. Licklider foresaw computers communicating in natural language, processing information rapidly and accurately, storing and retrieving large data sets, and adapting to user needs, thus predicting a networked global information system.

With the advent of generative AI, particularly chatbots that interact effectively with human users while possessing powerful computational and problem-solving capabilities, implementing centaurian models on a large scale has become more feasible. This discussion aims to provide guidelines for designing centaurian systems and compare these models with other approaches to hybrid intelligence.

The structure of this paper is as follows:

- **A Cognitive Architecture for Centaurian Models:** This section outlines the foundational framework for centaurian models, drawing on Herbert Simon's Design Science. It considers the evolution from Homo Sapiens to Homo Faber, culminating in the centaurian model that hybridizes human and artificial intelligence.
- **A System View of the Centaur Model:** This section categorizes centaurian systems into monotonic and non-monotonic models and uses distinctions between open and closed systems to provide a criterion for choosing between centauric evolution of a human-operated system or outright replacement by a machine-operated one.
- **Related Work:** This section critically compares centaurian models with three related theories and frameworks: the Theory of Extended Mind, Intellectology, and multi-agent approaches to hybrid intelligence, assessing their relationships, advantages, and limitations.
- **Conclusion:** The final section summarizes the insights from the study and outlines potential future research directions for centaurian models in hybrid intelligence.

In this paper, we use the terms "centaur", "centauric", and "centaurian" interchangeably, treating them as synonyms. This choice allows us flexibility in using these terms, reflecting common usage across various contexts and sources, including those beyond our authorship. We preferred to maintain this freedom of expression rather than impose a contrived uniformity in terminology.

2. A Cognitive Architecture for Centaurian Models

Centaurian models of hybrid intelligence are designed to harness the creative potential of integrating human and technological capabilities. This progression moves from Homo Sapiens, with innate cognitive abilities, to Homo Faber, who uses technology to transform the environment, ultimately evolving into Centaurus Faber, hybridizing human and artificial intelligence. Unlike models focused on human–AI feedback loops in specific machine learning contexts [3], centauric models emphasize broader systemic functionality, allowing the AI component to function as an opaque subsystem, focusing instead on behavioral integration. This perspective shifts the focus from machine training transparency to operational outcomes and functional coherence in system design, highlighting how centauric models prioritize seamless functionality in highly complex environments.

Our methodology is, in fact, grounded in Herbert Simon's seminal work, *The Sciences of the Artificial* [4], a cornerstone of Design Science. Simon's architecture delineates a tripartite structure: an external interface for world interaction, a coding mechanism for encoding environmental stimuli, and an internal processing system for creating artifacts. This framework is not simply theoretical but builds on centauric systems already in practice, as will be exemplified in the next section. Rather than proposing a purely experimental model, our approach formalizes the design of systems that hybridize human and artificial intelligence. The focus is on developing a methodology that can guide centaurian system creation, ensuring that design choices—such as replacing human functions with artificial ones or preserving human agency—are made judiciously. For example, as in the case of strategic games like chess, a complete replacement of human capabilities by AI may sometimes be the optimal choice, but this is contingent on the system's goals and functional requirements.

Simon's architecture can be considered intrinsically centaurian for several reasons. It boasts a flexible structure capable of incorporating components defined by their functionality rather than their origin, allowing the inclusion of biological components evolved through natural selection and synthetic components crafted through engineering efforts

like AI. Simon elucidates how these components, even when underpinned by biological hardware, function within the "artificial" realm, creating artifacts across diverse domains such as engineering, architecture, management, crafts, and art. The convergence between the natural and the synthetic in transformative activities is a further stage along this route.

This framework presents several pertinent challenges and inquiries central to this article. Notably, the autonomy of artificial components enables behaviors yielding less predictable outcomes. Simon's architecture distinguishes between interactions with the external world and internal processing. In a centauric perspective, we can conceptualize the components responsible for these two complementary activities as, respectively, "head" and "body". This suggests that a system's centaurian essence is defined by its components' functional division rather than their origin—natural or synthetic.

Thus, this framework reinforces that systems composed entirely of human or synthetic elements, or a mixture of both, inherently exhibit a centaurian character when engaged in transformative activities. This perspective underscores the intrinsic centaurian nature of systems that enhance human capabilities through artificial means, highlighting the seamless continuum between human and synthetic components within the centaurian model.

To illustrate the architecture's application, consider its diagrammatic representation in Figure 1 and the example of a civil engineer leveraging their creative and operational skills:

- **Leverage of the External Interface:** The engineer employs artificial sensors—such as surveying tools, aerial imagery, and ground-penetrating radar—to collect data about the project environment and define specific requirements.
- **Coding of Environmental Stimuli:** The engineer translates the collected data into project drawings, models, and diagrams, assisting in visualizing the project and informing decisions regarding materials, design, and construction methodologies.
- **Internal Processing System for Artifact Creation:** The engineer interprets the information using reasoning, problem-solving, and decision-making techniques to evaluate the feasibility and cost of various design options and identify the most effective strategy.
- **Feedback Loop:** The engineer gathers feedback to refine representations and procedures as the project advances. For instance, if unexpected challenges emerge during construction, new data is collected and analyzed to address these issues effectively.
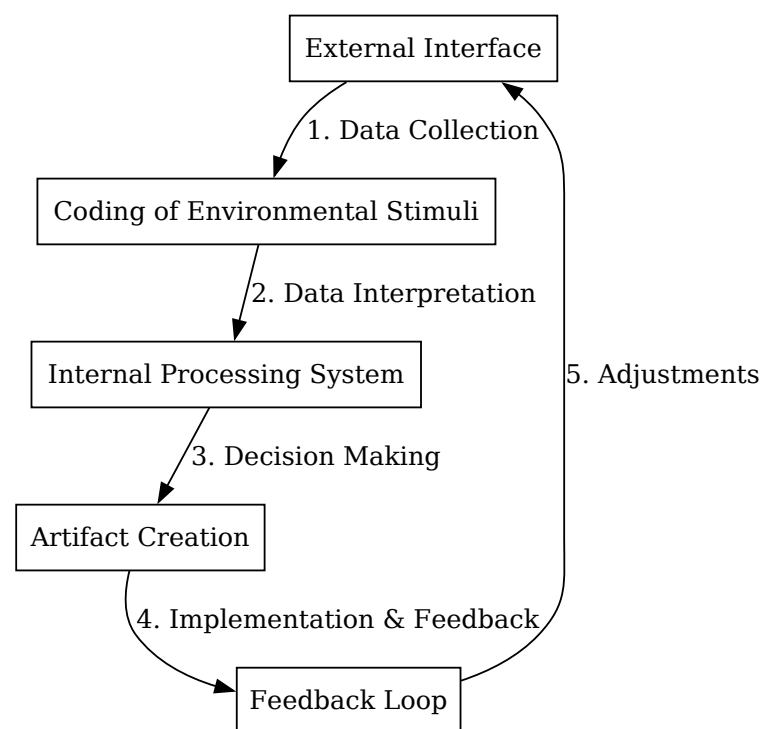

Figure 1. Simon's Cognitive Architecture.

The inherent modularity of Simon's architecture is evident in the diverse outcomes it enables: engineers from various traditions, applying the same process, can devise distinct solutions to identical problems. This flexibility and adaptability are essential for creative problem-solving and pave the way for extending this approach to centaurian systems. This extension involves substituting, whether totally or partially, one or more components initially derived from human training and biological foundations with synthetic counterparts. Operating on non-biological (silicon-based) hardware, these synthetic components derive their problem-solving capabilities from algorithmic implementation and execution.

3. A System View of the Centaur Model

Understanding centaurian models as systems is crucial for effectively designing and adopting these hybrid intelligence frameworks. Centaurian models function systemically due to their multi-component nature, whether composed of human elements or synthetic components. This section delves into the core characteristics of these systems, focusing on two key attributes: their classification as monotonic or non-monotonic and the distinction between open and closed systems. This analysis provides a foundation for designing centaurian systems that can be seamlessly integrated into various contexts.

3.1. Monotonic and Non-Monotonic Centaur Systems

A critical aspect of centaurian systems is how they integrate synthetic capabilities to either preserve or enhance human capabilities, classifying them into two categories: *monotonic* and *non-monotonic* systems.

Monotonic centaurian systems ensure synthetic components consistently enhance or maintain existing human capabilities, offering predictability and stability. They are ideal for established contexts where maintaining semantic coherence and system integrity is crucial.

In contrast, non-monotonic centaurian systems may introduce significant changes and innovations that can disrupt existing processes. These systems are valuable in scenarios where radical innovation can bring substantial benefits, even if it means overhauling current practices and potentially introducing some instability.

To better understand and classify these systems, we draw upon and adapt a well-established principle from software evolution: the *Liskov Substitution Principle* (LSP) [5]. Widely recognized in software engineering, particularly within object-oriented programming languages such as C++ and Java, the LSP stipulates that a subclass replacing its parent class must not alter the application's functionality or integrity. This principle ensures that objects of the superclass can be seamlessly substituted with objects of the subclass, thereby maintaining the program's correctness and expected behavior.

In centaurian models, the principles underlying LSP for object-oriented programming offer valuable insights for ensuring seamless integration between human and artificial intelligence. Specifically, the following aspects of LSP can be adeptly adapted to centaurian systems:

- **Method Parameters:** In software development, LSP requires a subclass to override superclass methods without altering expected behavior. This ensures system integrity by mandating that the subclass's methods accept inputs and produce outputs consistent with those of the superclass. This consistency ensures that integrating artificial components into human-centric workflows enhances rather than disrupts existing processes. It underscores the importance of compatibility and coherence in merging human and machine capabilities, especially in mission-critical domains where semantic coherence and operational stability are essential.
- **Return Values:** Similarly, the outputs generated by a synthetic component within a monotonic centaurian system must conform to the expectations set by the human capabilities they aim to replicate or augment. This ensures that subsequent processes or decisions relying on these outputs are valid and effective. This mirrors the software engineering principle where the return value of a method in a subclass must meet the

expectations set by the method in the superclass. Such alignment is fundamental to maintaining the system's overall integrity and functionality.

3.1.1. Monotonic Centaurian Systems

Monotonic centaurian systems effectively embody the essence of the Liskov substitution principle (LSP) in human-AI collaboration. These systems are designed to ensure that any synthetic augmentation or substitution of human capabilities matches and potentially surpasses the functional capacity of the original human component while avoiding discrepancies or incompatibilities. The concept of "monotonicity" here pertains to the consistent preservation and enhancement of functionality, ensuring a non-decreasing trajectory in the system's capabilities as it transitions from human to synthetic components. Monotonic centaurian systems are characterized by the following attributes:

- **Seamless Integration:** Synthetic components are meticulously designed to integrate with existing human cognitive processes, ensuring the retention of functionality and semantic coherence. This integration preserves the natural flow of operations without compromising system coherence or functionality.
- **Functional Extension:** These systems go beyond simple replacement strategies. Synthetic components are engineered to introduce additional capabilities or efficiencies, thereby broadening the scope and enhancing the system's overall functionality.
- **Preservation of System Integrity:** By adhering to established operational parameters and practices, synthetic components are integrated to maintain or elevate the system's integrity and performance. This ensures that the foundational qualities of the system are not only preserved but are potentially enhanced.

To further illustrate this, we now provide a detailed monotonic centauric system case study, focusing on a cognitive trading system. After this, additional use cases will be described more synthetically.

Cognitive Trading System

The cognitive trading system presented in [6], a chapter in the book [7], exemplifies a monotonic centauric system within the financial sector and indeed significantly contributed to the motivation and relevant background knowledge for the foundational approach to centaur models developed here. This system integrates heuristic trading strategies—such as those found in technical analysis [8]—with neuro-evolutionary networks that dynamically adapt and optimize trading strategies in real-time [9]. The human trader defines the initial strategy based on heuristic knowledge and market conditions. At the same time, the AI components enhance these strategies through machine learning, enabling real-time adaptation to changing market environments.

The system respects the LSP, thus ensuring that AI-driven components augment human decision-making without disrupting the underlying framework. As the AI refines trading strategies, it does so in a manner that preserves the functional integrity of the original human strategy, thus ensuring monotonicity. This centauric system enhances the overall system's functionality by enabling human traders to leverage advanced AI techniques while maintaining high-level strategic control.

The neuro-evolutionary networks employed within this system evolve and optimize artificial neural networks to ensure efficient strategy execution without introducing non-monotonic behavior. This ensures that, as additional AI components are incorporated, the system's performance improves without sacrificing reliability or coherence. In this sense, the cognitive trading system is a prime example of centauric models enhancing human cognitive capacities without compromising operational integrity (see Figure 2).
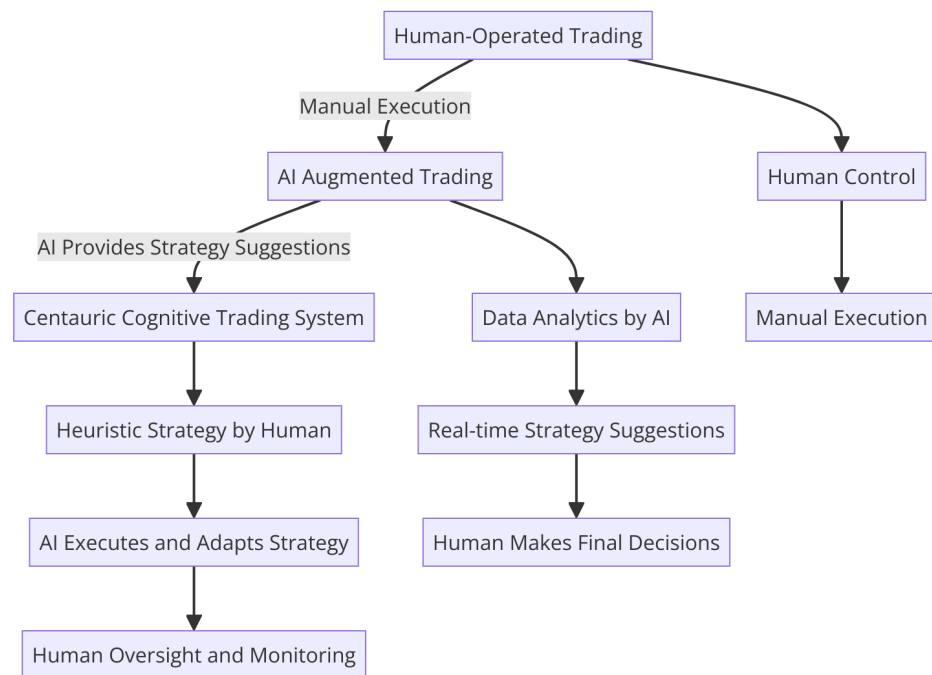
Figure 2. The Evolution of the Cognitive Trading System from Human-Operated Trading to AI-Augmented Trading to Centauric Cognitive Trading System.

Other Use Cases

In addition to the cognitive trading system, several other use cases exemplify monotonic centaurian systems:

1. **Automated Call Center Operators**
   a. **Substitutes:** External Interface and Internal Processing Component
   b. **Explanation:** Automated call center operators, powered by natural language processing (NLP) systems, are prime examples of monotonic centaurian systems. These systems receive and interpret verbal or written input from users, generating appropriate responses to facilitate enhanced interactions between humans and digital environments. By substituting both the external interface and internal processing components with NLP technologies, these systems maintain the functional capacity of human operators while potentially extending their capabilities. This substitution allows seamless integration into existing call center workflows, improving the system's efficiency and effectiveness without compromising semantic coherence or operational integrity. Furthermore, humans can remain in the loop by providing second-level or higher-level services. While AI replaces human functions at the lowest level, humans retain control, enabling oversight and intervention when complex or nuanced interactions arise. This hierarchical integration ensures that human expertise is utilized where most impactful, enhancing the overall service quality [10].

2. **Semi-Autonomous Drones**
   a. **Substitutes:** External Interface and Internal Processing Component
   b. **Explanation:** Semi-autonomous drones, guided by human pilots but capable of performing certain actions autonomously, exemplify monotonic centaurian systems. These drones use advanced sensors and AI to perceive their environment, make informed decisions, and execute specific tasks. However, human operators remain actively involved, overseeing missions and intervening as necessary. This integration allows drones to enhance human capabilities, particularly in environments where direct human participation is risky or impractical. By

blending AI-driven processing with human oversight, semi-autonomous drones maintain system functionality and effectiveness while ensuring critical decision-making remains under human control. This collaborative approach raises important questions about responsibility and decision-making in scenarios such as military operations, where the actions of autonomous or semi-autonomous systems can have significant consequences. Finding the right balance between human oversight and machine autonomy in these systems highlights the ethical and practical considerations central to deploying centaurian models [11].

3. **Cognitive Assistants**

   a. **Substitutes:** External Interface and Internal Processing Component
   b. **Explanation:** Cognitive assistants exemplify the seamless integration of artificial intelligence in enhancing human cognitive functions. By personalizing interactions and recommendations through sophisticated processing of user data, these systems act as the interface for data collection and as mechanisms for adaptive learning and decision support. Cognitive assistants extend and augment human capabilities in processing information and making informed decisions, aligning perfectly with the principles of monotonic centaurian systems by maintaining and enhancing the system's overall functionality without introducing semantic discrepancies or operational incompatibilities [12].

3.1.2. Non-Monotonic Centaurian Systems

Non-monotonic centaurian systems offer higher innovative potential but do not ensure the preservation of existing functionalities and may disrupt the system's semantic coherence. Risks associated with non-monotonic centaurian systems include:

- **Disruptive Integration:** Inconsistencies or conflicts within pre-existing cognitive or operational frameworks, posing challenges to system integrity.
- **Variable Functionality:** Synthetic components in non-monotonic systems exhibit fluctuating functionality, which may not consistently align with the capabilities they aim to replace. This variability can lead to potential reductions in system performance.
- **Semantic Incoherence:** Risk of disrupting the semantic coherence inherent in human cognitive processes, potentially resulting in misinterpretations or errors that deviate from established norms and practices.

The potential for disruptive innovation in non-monotonic systems is notably higher than in monotonic ones. However, their risks render them less suitable for domains where stability, predictability, and semantic coherence are paramount.

Systems initially designed as monotonic can evolve into non-monotonic entities over time. Such transitions underscore the inherent fluidity and adaptability of centaurian systems, revealing a spectrum that spans from strict predictability to innovation and unpredictability. Maintaining oversight and control over the development trajectory of centaurian systems is crucial. Decision-makers face a pivotal choice: embrace the uncertainties and opportunities of non-monotonicity or adhere to monotonic systems' safer, more predictable path. This choice is not merely technical but strategic, influencing the direction of innovation and the potential for realizing untapped possibilities within the centaurian framework.

3.1.3. Centaurian Evolution

The evolution of a centaurian system can be visualized as a geometric translation, where newly introduced components assume the functions previously performed by their predecessors. For instance, the progression of an NLP system, as depicted in Figure 3, showcases a shift towards artificial components for both internal processing and external interfacing. The system can retain its monotonic character if changes adhere to the Liskov substitution principle (LSP), which is easily enforceable in scenarios with highly structured human interactions, such as those encountered in call center environments.
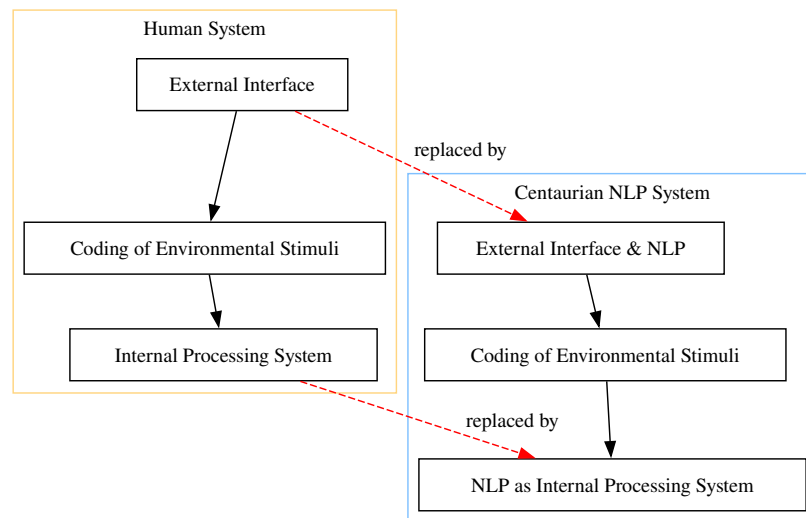
Figure 3. Evolution of a Centaur NLP System.

However, these dynamics may change significantly when delving into domains inherently defined by freedom and creativity, such as art and design. Introducing synthetic components for interpreting external stimuli and generating creative outputs introduces complexity and unpredictability. This shift may mark a departure from monotonicity, transitioning the system into a domain where creative and interpretive processes are no longer bound by linear enhancements of human capabilities but are propelled by the autonomous generative potential of artificial intelligence.

A leading example of this evolution is the utilization of text-to-image (TTI) platforms like DALL-E or stable diffusion, as illustrated in Figure 4. Initially, these platforms serve internal processing roles, guided by instructions from a human operator who ensures the alignment of synthetic outputs with external stimuli, thereby maintaining the system's monotonic nature. The evolution accelerates when the interface with the external world becomes artificial, such as by incorporating an artificial image-to-text decoder. This transition significantly alters the interpretation of external reality, moving away from the subjective human perception crucial to artistic production. At this juncture, we venture into the richly potential yet unpredictable domain of non-monotonicity, characterized by its capacity for innovation and the opening of new opportunities.

Even when exploring these uncharted waters—rich with potential discoveries but not without risks—it is possible to design guided and structured paths that mitigate risk while fostering creativity. An example of such a path is illustrated in Figure 5, the "Creativity Loop" of CreatiChain [13]. In this framework, AI-powered automation generates images and prompts from relevant ontologies and domain lexicons. Initially unbridled, this process later returns control to the human creator, who can refine and evolve the outputs, thus blending AI's unpredictability with human creative oversight.

In its initial phase, the CreatiChain Creativity Loop leverages external interfaces and internal processes to generate creative prompts across diverse domains. This phase is marked by the typical unpredictability of AI behavior, a hallmark of non-monotonic centaurian systems. Subsequently, human creators regain control, selecting, modifying, and evolving AI-generated outputs to align with their creative vision, showcasing the transformative potential of AI in creative industries.

Similarly, the study described in [14] explores the synergy between vision-language pretraining and AI-based text generation (through ChatGPT) for artwork captioning, providing another instance of centauric integration. The researchers introduce a dataset for artwork captioning, refined through prompt engineering and AI-generated visual descriptions. This approach further illustrates the transformative potential of AI in creative

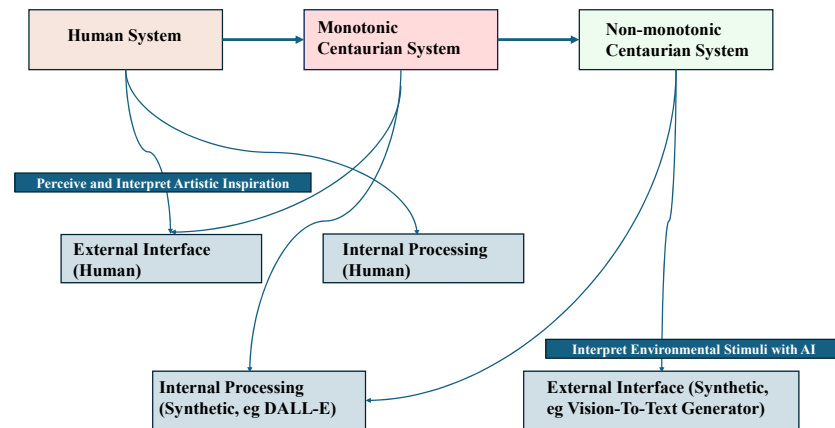processes, enabling human creators to refine and evolve AI-generated outputs to fit their creative vision.

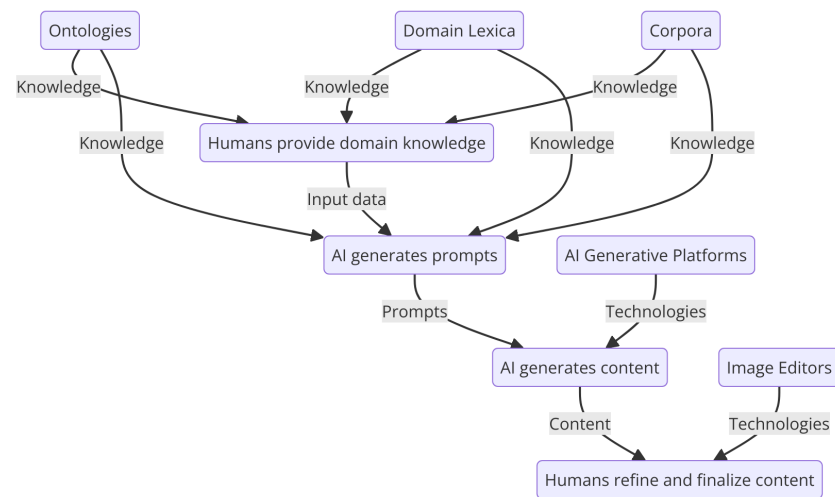**Figure 4.** From Monotonic to Non-monotonic.

Figure 5. CreatiChain Creativity Loop.

Both frameworks described above provide instances of effectively integrating non-monotonic systems within a creative context, enabling users to expand their creative horizons and explore previously untapped territories. Fostering such a dynamic interplay between human insight and artificial intelligence significantly expands creative freedom and innovation, underscoring the transformative potential of collaborative creativity in the digital age.

3.1.4. Monotonic or Non-Monotonic?

Applying non-monotonic systems to structured domains such as healthcare, law, and infrastructure requires careful consideration. These sectors demand predictability, reliability, and transparency, which non-monotonic systems may challenge with their inherent unpredictability. The risks include accountability issues, bias, and potential errors with significant real-world consequences.

Monotonic systems may be preferred in regulated domains to balance innovation and reliability. These systems augment human capabilities predictably and transparently, ensuring safety, accountability, and ethical integrity. In healthcare, for example, ensuring trustworthiness and ethical alignment becomes crucial. It may require going beyond the simple, functional principle given by the Liskov Substitution Principle by also resorting to methodologies to assess AI systems' impact and maintain the desired ethical alignment. Consider, for instance, the following two case studies, described, respectively, in

articles [15,16] and published in a special journal issue devoted to the collaboration between human and artificial intelligence in medical practice [17]:

- **Cognitive Assistants in Cardiac Arrest Detection:** This AI system enhances decision-making by providing real-time alerts during emergency calls. Its monotonicity is maintained by supporting dispatchers without diminishing their autonomy. Ethical concerns include alert fatigue, bias, and transparency.
- **Cognitive Assistants in Dermatology:** This AI system aids dermatologists in diagnosing skin lesions by offering data-driven insights while preserving human oversight. Monotonicity is ensured by complementing, not replacing, human judgment. Ethical challenges include algorithmic bias and transparency, which must be addressed to ensure the system is trustworthy and ethical.

In both cases, a methodology (Z-Inspection® [18]) was used to ensure the required ethical standards. It aligns well with centauric system design principles by promoting human oversight, transparency, and ethical considerations.

The legal domain offers another promising field for applying monotonic centaurian systems, given the complexity of legal reasoning and the stringent requirements for transparency and accountability. Articles [19,20] provide in-depth discussions and case studies that offer valuable insights for building such systems in legal practice. These studies illustrate how models integrating human judgment with AI assistance should maintain the necessary balance, underscoring the importance of predictability and reliability. They emphasize that in domains like law, AI should be designed to complement human decision-making rather than replace it, aligning with the principles of monotonic centaurian system design.

In contrast, non-monotonic systems excel in creative fields where outcomes cannot be easily quantified. These systems' adaptability and unpredictability make them ideal for generating innovative and novel results in art, design, and literature. While they pose risks in regulated environments, their potential to revolutionize creative industries is substantial, raising them to unprecedented levels of innovation and artistic expression.

3.2. Open vs. Closed Systems in Centaurian Integration

Building on the exploration of monotonic and non-monotonic centaurian systems and their impact on human decision-making processes, we now confront a more fundamental question: How do we choose between man-machine integration, that is, actual centauric systems, and the outright replacement of processes previously managed by humans? This inquiry is crucial for understanding the limitations of centaurian systems in domains like chess and their potential in areas such as art, design, and organizational decision-making. A key distinction in addressing this question is understanding systems as either "open" or "closed".

While rooted in physics, the concept of open and closed systems extends significantly into socio-economic and information technology domains. In physics, closed systems are defined by their inability to exchange matter with their surroundings, though they can exchange energy. Conversely, open systems are characterized by their ability to exchange both energy and matter with their environment. This fundamental distinction provides a lens through which we can examine human and machine intelligence integration.

3.2.1. Characterizing Open and Closed Systems

Visionary computer scientist Carl Hewitt has significantly contributed to our comprehension of systems as open or closed within information processing and decision-making domains. His seminal work on concurrent computation and the Actor Model establishes that organizations operate as open systems, hallmarked by due process and a dynamic exchange of information [21]. This insight is pivotal, suggesting that open systems are predisposed to evolve in a centauric direction—characterized by an internal processing capacity bolstered by an exchange of information with the external environment. This exchange is distributed across multiple elements, resonating with Herbert Simon's cognitive

architecture, comprising an external interface and internal processing capabilities. These components are foundational to the definition and characterization of centaurian systems.

In parallel, the philosopher and sociologist Edgar Morin developed a complexity paradigm that imparts a multidimensional perspective on systems. Morin's framework diverges from conventional linear models, advocating for an acknowledgment of complexity, recursive feedback loops, and the "unitas multiplex" principle. This principle, central to Morin's philosophy, emphasizes the intrinsic unity within diversity. It posits that systems are not mere assemblies of disparate elements but rather cohesive entities displaying a rich mosaic of interrelations and interdependencies [22].

Morin's concept of "unitas multiplex" posits that every system embodies a symbiosis of unity and multiplicity. According to Morin, systems are intricate unities encompassing multiple dimensions and scales of interaction. They are recursive, signifying their self-referential nature and capacity for self-regulation via feedback loops. Furthermore, these systems are adaptive, exhibiting resilience and the ability to evolve in response to both internal and external stimuli.

Morin's interpretation of "unitas multiplex" suggests that systems exhibiting such traits are prime candidates for evolution in a centauric direction, which, in essence, is a manifestation of "unitas multiplex".

Ultimately, although Hewitt and Morin approach the study of systems from distinct academic disciplines, their theories converge on open systems and the delicate equilibrium between unity and diversity. Their insights guide us in discerning which systems are amenable to evolution in a centauric direction and which may be more effectively addressed in reductionist terms—that is, through the complete substitution of the human component with a more efficient and high-performance artificial counterpart.

3.2.2. Explaining and Predicting Reductionism

Centaur Chess, once heralded as a beacon of innovation, represents the collaboration between human intellect and the computational might of chess engines. Initially perceived as an effective blend that could elevate the human element to the zenith of chess mastery, Centaur Chess has encountered insurmountable challenges. These obstacles have precipitated its decline in prominence, signaling a notable retreat from the aspiration to reestablish human preeminence in chess. An examination of Centaur Chess's descent, through the interpretive lenses of Carl Hewitt's open and closed systems and Edgar Morin's "Unitas Multiplex", elucidates the efficacy of these theoretical constructs in guiding the preferability of a reductionist approach for system evolution.

- **Open and Closed Systems** (Hewitt): Hewitt's delineation of open systems underscores the vitality of information exchange and interoperability. Within the milieu of Centaur Chess, the human-computer amalgam epitomizes an open system, synergizing human strategic acumen with the algorithmic prowess of chess engines. Yet, the viability of this symbiosis hinges on the exchange's caliber and pertinence. With the relentless progression of chess engines, human-provided insights may no longer substantively augment or refine the engine's computations, culminating in a waning efficacy of human involvement. This observation intimates that the open system paradigm of Centaur Chess may have ceded its advantage to the more streamlined, closed system of autonomous chess engines.
- **Unitas Multiplex** (Morin): Morin's "Unitas Multiplex" concept articulates the coalescence within diversity and the intricate interplay across a system's multifarious dimensions. The Centaur Chess construct ostensibly embodies this multiplicity by integrating human and computational elements. Nevertheless, the purported unity of this composite system is beleaguered by the computational component's overwhelming analytical ascendancy. The confluence of human and machine contributions does not inherently forge a more potent unified entity; rather, it may engender complexity devoid of commensurate efficacy.

In extrapolating these theoretical frameworks to Centaur Chess, we discern that:

- **Centauric Approach:** Pursuing a centauric system in chess, aspiring to meld human and machine intelligence, may not represent the optimal strategy. The intrinsic requirement for efficacious information exchange within an open system is unfulfilled when the computational element can function with superior autonomy.
- **Reductionist Approach:** Conversely, a reductionist approach, eschewing the human factor in favor of exclusive machine reliance, is congruent with AI's prevailing trajectory in chess. The self-contained system of a chess engine, devoid of human intercession, has demonstrably outperformed and achieved unparalleled levels of play.

In summation, while the inception of Centaur Chess was marked by ingenuity and potential, the pragmatic application of Hewitt's and Morin's theoretical perspectives intimates that an evolutionary trajectory towards a centauric model is less tenable than a reductionist paradigm capitalizing on AI's unbridled capabilities. The solitary AI chess engine exemplifies a closed system, self-sufficient and unencumbered by the heterogeneous inputs emblematic of "Unitas Multiplex". This reductionist stance has indeed emerged as the preeminent pathway to the apex of chess proficiency.

3.2.3. Advocating for Centauric Systems in Art

The domain of art, unlike strategy games, is inherently an open system that thrives on interaction and adaptability. This makes it particularly suitable for centauric approaches, where AI's computational capabilities enhance human intuition and creativity. By examining the principles of Carl Hewitt's open systems and Edgar Morin's "Unitas Multiplex", we can elucidate why the centauric model is not only feasible but also highly advantageous in the artistic domain.

- **Open Systems and Centauric Art** (Hewitt): Carl Hewitt's framework emphasizes the dynamic exchange of information and the importance of interoperability in open systems. Art continually interacts with its environment as an open system, incorporating diverse stimuli into the creative process. This ongoing interaction aligns perfectly with the centauric model, where AI tools like DALL-E or other generative platforms serve as creative partners, augmenting the artist's capabilities rather than replacing them. The human artist provides the conceptual framework and interpretive insight, while the AI contributes novel variations and enhances the execution of artistic ideas.
- **Unitas Multiplex in Art** (Morin): Edgar Morin's principle of "Unitas Multiplex" encapsulates the unity within diversity, which is intrinsic to the artistic process. Art thrives on the fusion of various influences, styles, and techniques, embodying a complex interplay of internal and external elements. The centauric approach, which integrates human creativity with AI's generative potential, epitomizes this principle by creating a synergistic partnership that enriches the creative landscape. AI's ability to generate new ideas and iterate on artistic concepts complements the human artist's vision, leading to a cohesive and innovative artistic expression.

By extrapolating these theoretical frameworks to the realm of art, we discern that:

- **Centauric Approach:** In the domain of art, the centauric approach offers a powerful model for collaboration between humans and machine intelligence. The intrinsic openness of the artistic process benefits from the continuous exchange of ideas and feedback between the artist and AI. This collaboration enhances creativity, allowing for exploring new artistic territories and creating works that neither humans nor machines can achieve independently.
- **Sustaining Human Involvement:** Unlike the closed systems of strategy games, where AI can fully substitute human players, the open system of art necessitates ongoing human involvement. The centauric model ensures that human intuition, emotion, and cultural context remain integral to the artistic process, with AI serving as a tool that expands the artist's creative toolkit.

In practical terms, the centauric approach in art can be illustrated through various examples. For instance, generative platforms like DALL-E allow artists to input textual

descriptions and receive visual outputs as inspiration or starting points for further development. This process aligns with Simon's cognitive architecture, where the human artist (sensor interface) gathers information and transforms it into prompts for the AI (internal processor), resulting in a collaborative creation process.

The two evolutionary directions of chess and art are depicted and compared in Figures 6 and 7, respectively. Figure 6 illustrates a closed system that has strongly favored a reductionist trajectory in the world of chess. Within this system, humans are relegated to the role of 'second-class' players, clearly separated by an unbridgeable gap in skill and playing strength from 'first-class' artificial players. In this defined ecosystem, technology providers, predominantly humans, still play a significant role, while human chess enthusiasts and amateurs maintain a collateral interest in both the A-league (AI players) and B-league (human players), as indicated by the dotted lines. This ecosystem is otherwise isolated from further interactions with the broader world, consistent with its inherently closed nature.
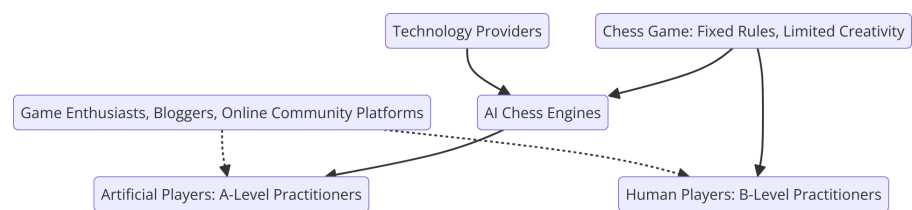
Figure 6. Evolution of Chess Systems: Closed/Reductionist Approach.
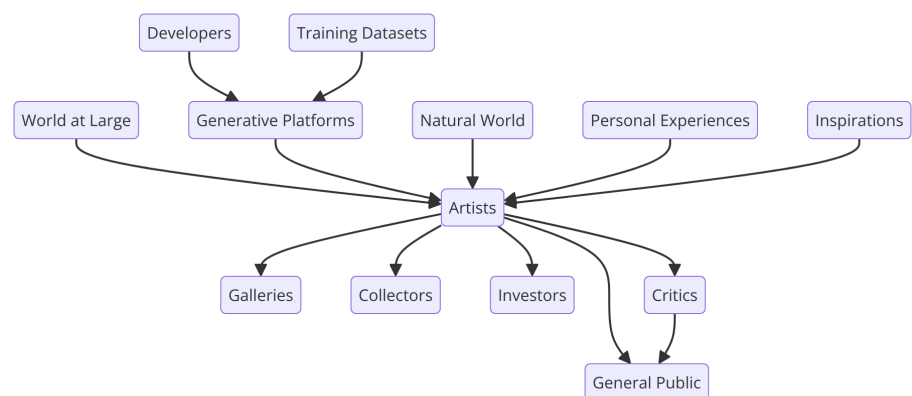
Figure 7. Evolution of Art Systems: Open/Centauric Approach.

In contrast, Figure 7 unequivocally illustrates the openness of the art world, where artists remain the central hub, processing and re-elaborating a diverse array of stimuli and inputs. This process is now centaurically augmented through the integration of generative platforms, along with the actors and components that support them. The artists' creative activity continuously feeds back into the external world, influencing and transforming it in an ongoing, dynamic cycle.

3.3. Leveraging the Monotonic/Non-Monotonic and Open/Closed Distinctions for Design

We have introduced two critical distinctions: monotonic and non-monotonic centauric systems and open and closed systems. Monotonic and non-monotonic systems represent different evolutionary paths for centaurian systems. Open and closed systems, on the other hand, help determine whether to pursue hybrid centauric upgrades or opt for a complete replacement of the human component with artificial counterparts. Open systems are conducive to hybridization, while closed systems often lend themselves to full replacement by artificial intelligence.

Notably, highly creative contexts such as art are paradigmatic in terms of both non-monotonic innovation and the suitability of centaurian hybridization. Furthermore, even

when centaurian systems evolve to feature synthetic components predominantly, they remain distinct from fully artificial systems due to their inherent heterogeneity and functional diversity. This distinction underscores the systemic openness that justifies their evolution in a centauric direction. The whole domain of art can be viewed as naturally leading to open creative systems of various kinds, as detailed in the book *Centaur Art* [23], where the dynamic interplay between human creativity and artificial ingenuity is thoroughly explored, illustrating how generative AI and hybrid systems can transform artistic practices. This book highlights historical and contemporary examples and underscores the potential for non-monotonic innovation within the art world, providing a comprehensive framework for understanding and harnessing the creative synergy between humans and machines.

4. Related Work

We address the comparison and discussion of the related work by distributing it across three main themes: the Theory of Extended Mind, artificial general intelligence and Intellectology, and multiagent approaches.

4.1. The Extended Mind

The discussion of related work naturally begins with a comparison to the Theory of the Extended Mind, a framework introduced by philosophers Andy Clark and David Chalmers. In their seminal work [24], they propose a dualistic structure of the human mind, encompassing both an internal component—rooted in the brain and neurons housed within the skull—and potential external extensions, such as computational tools ranging from simple devices like pocket calculators to increasingly sophisticated platforms. While comparing centauric systems and the Theory of the Extended Mind is inevitable, given their shared focus on hybrid human–machine contexts, it is crucial to acknowledge a fundamental difference in intent between the two frameworks.

Centauric models are primarily concerned with defining a practical methodology for designing and implementing hybrid human-machine systems. In contrast, the Theory of the Extended Mind aims to offer a framework for describing and explaining the human mind. This difference in purpose leads to the conclusion that centauric models are fundamentally agnostic with respect to the Theory of the Extended Mind. They are compatible with the Theory of the Extended Mind but can also align with more traditional views that confine the mind strictly to the brain. This flexibility underscores that centauric models do not necessarily depend on any particular philosophical stance about the nature of the mind. From a strictly neurological perspective, where the mind is viewed as a byproduct of brain activity, centauric systems can still be effectively designed and implemented, with the focus remaining on practical integration rather than philosophical definitions.

To fully grasp why this is the case, let us briefly overview the arguments for and against the Theory of the Extended Mind. The main arguments in favor can be summed up as follows:

1. **Parity Principle:** Clark and Chalmers introduce the "Parity Principle", suggesting that if an external tool or resource functions in a way that we would consider cognitive if it occurred within the brain, then it should be considered part of the cognitive process. For example, if a person uses a notebook to store information just as another might use their biological memory, the notebook becomes part of their extended mind.
2. **Empirical Support:** Various empirical studies, such as Esther Thelen and Linda Smith's work on infant motor development [25], as well as vision research [26] and behavioral robotics [27], demonstrate how external factors significantly influence cognitive processes. These studies highlight how interactions with the environment can alter or enhance cognitive functions, supporting the idea that the mind can extend beyond the brain.
3. **Functional Integration:** The theory emphasizes integrating internal and external elements seamlessly in cognitive processes. For example, in his book *Being There: Putting Brain, Body, and World Together Again*, Clark [28] illustrates how Scrabble

players manipulate tiles to stimulate word recognition, demonstrating the interaction between physical actions and cognitive processes. Similarly, the concept of *iterative sketching*, as characterized by 20th-century art historian Ernst Gombrich [29], though developed independently, can be used to support the view that external tools are integral components of cognition.

Conversely, these are the main critiques and counterarguments:

1. **Constitutive vs. Causal Relationships:** Critics like Fred Adams and Kenneth Aizawa [30] argue against the constitutive claim of the extended mind theory, suggesting that external tools merely cause cognitive processes rather than constitute them. They warn against the "fallacy of constitution", where causal relationships are incorrectly interpreted as constitutive.
2. **Cognitive Boundaries:** Robert Rupert [31] and other critics maintain that the biological organism should bound cognitive processes. They argue that extending cognition to include external elements could undermine the coherence of cognitive science, which relies on well-defined, organism-bound systems.
3. **Intrinsic Intentionality:** Critics such as Adams and Aizawa also contend that genuine cognitive processes must exhibit intrinsic intentionality—an inherent mental content. External tools, they argue, do not possess this quality and thus cannot be considered part of the mind.

Centauric models, however, can be applied to both the Theory of the Extended Mind and its opposite, the Theory of the Embedded Mind, which maintains the traditional view that cognition is confined to the brain without requiring structural changes to the system itself. For example, consider the cognitive trading system case study, [6], which was used here as a paradigmatic example of a monotonic centaurian system:

*   When interpreted through the extended mind (EM) lens, the cognitive trading system extends the trader's mind by integrating AI's data-processing capabilities into the trader's cognitive workflow. Here, the AI system is considered part of the trader's extended cognitive apparatus, external but essential for decision-making. The centauric system's functional coherence and interdependency align with the EM framework by seamlessly extending the cognitive processes.
*   Conversely, within the embedded mind (EMb) view, the cognitive trading system remains a sophisticated tool external to the mind. The trader remains the cognitive agent, using the AI as an external resource to enhance decision-making. Even here, the centauric model holds because the focus is on robust functional collaboration rather than philosophical interpretation, showing how centauric systems are orthogonal to both cognitive theories.

This example demonstrates that centauric models are not philosophically bound to a particular theory of cognition. Whether the AI system is viewed as an extension of the human mind (extended mind) or a tool (embedded mind), the centauric system's key goal remains functional integration and seamless human–AI collaboration.

When considering the relationship between the Theory of the Extended Mind and centaur models, it is essential to frame this in the context of advanced artificial intelligence. Recent developments have introduced significantly more sophisticated possibilities for external extensions than those addressed in the original formulation of the extended mind. While the original theory naturally focused on computational extensions, such as pocket calculators and personal computers, it did not account for extensions with autonomous creative capacities that replace primarily human tasks, from autonomous driving to artistic creativity.

This enhancement and autonomy of artificial capacities introduce new challenges to the Theory of the Extended Mind. Specifically, it raises the question of whether such autonomous capacities can still be considered external extensions of the human mind or if they represent fully independent synthetic minds. Centaur models, allowing for the

functional integration of natural and synthetic cognitive capacities, maintain a less stringent requirement for artificial extensions.

Centaur systems can thus operate independently of the Theory of the Extended Mind. While they may align with refined variants of the theory, they also remain neutral, being compatible with more traditional views of the mind that align with objections to the extended mind. Some recent arguments in favor of the extended mind [32] highlight how the theory is naturally oriented to the cognitive dimension of Homo Faber—which inherently relates to external tools. Nevertheless, these are more socio-cultural arguments than empirically evidence-based.

The artistic domain is a specific area of interest, especially considering the arguments supporting the Extended Mind Theory regarding iterative sketching. Integrating AI-driven creativity with human artistic processes provides a compelling case for exploring the intersection of centaur systems and the extended mind.

4.2. Artificial General Intelligence and Intellectology

Although not as scientifically and academically established as the Theory of the Extended Mind, the concept of Intellectology is certainly relevant in relation to centauric models, with which it shares some goals, albeit remaining fundamentally different in its foundational and methodological approaches. Like the Theory of the Extended Mind, Intellectology seeks to broaden the definition and conceptualization of the mind beyond its traditional identification with its cranial location in the human body. However, the starting points of these two frameworks are markedly different: the Theory of the Extended Mind roots the mind in the human brain and external components, whereas Intellectology starts from a general, mathematical-computational definition of the mind. In this framework, the human mind is just one instance among a broader and arguably more significant space of minds, primarily populated by artificial intelligences and super-intelligences.

This foundational difference is also reflected in the distinct disciplinary backgrounds of the two theories of mind. The Theory of the Extended Mind emerges from the philosophy of mind and neuro-cognitive sciences. In contrast, Intellectology arises from a diverse milieu of computer scientists, technology entrepreneurs, bloggers, and journalists who are both fascinated by the potential of developing artificial general intelligence (AGI) and deeply concerned about the existential risks it might entail. This socio-cultural context spans a broad spectrum—from the transhumanist optimism of Ray Kurzweil, who envisions, as outlined in his bestseller *The Singularity Is Near* [33], a future where technological hyper-acceleration leads to the fusion of humans and machines, ultimately transcending the limitations of biological mortality, to theorists of existential risk like Nick Bostrom, author of *Superintelligence* [34], and Roman Yampolskiy. Both Bostrom and Yampolskiy warn of scenarios in which malevolent and hostile artificial super-intelligences could enslave or even exterminate the human species. However, Yampolskiy stands out for his systematic approach to these concerns through Intellectology, where one of his primary motivations is to rigorously define a taxonomy of intelligences and super-intelligences as part of a strategic effort to identify and address existential risks to humanity by clearly delineating the characteristics of potential hostile entities.

Yampolskiy's intent to create a comprehensive framework for systematically addressing the notion of mind—encompassing both natural and artificial minds of varying capacities and characteristics—is certainly commendable, regardless of his more or less pessimistic outlook on the future of artificial intelligence. From a purely investigative standpoint, and setting aside the differing practical motivations, Yampolskiy's goal of systematically treating these interactions aligns with the approach proposed here for centaur models. However, at least in its current form, his approach appears to be affected by significant challenges.

Following an initial conceptual introduction to Intellectology in a 2015 article [35], Yampolskiy's 2023 work [36] attempts a more formal and systematic treatment of the field. One of the main challenges in this approach is the definition of minds as binary strings. While this abstract mathematical representation has the advantage of positioning all minds

within a defined space, it raises significant concerns about differentiating between binary strings that represent minds and those that do not. The article does not clearly explain the criteria distinguishing "minds" from non-minds within this formalism. This ambiguity risks overgeneralizing the concept of a mind, diluting its meaningfulness, and leading to a definition that is too broad to be practical. Without specific constraints or additional structure, the definition appears too vague or incomplete to be useful, particularly in fields such as AI development, where precise and operationalizable definitions are essential.

Moreover, after introducing a formal, mathematical definition of the mind—suggesting a framework grounded in precision and quantitative analysis—the article's transition to conceptual and qualitative taxonomies lacks a clear connection to this formalism. This disconnect results in inconsistencies or a lack of coherence between how minds are defined and categorized, making the taxonomy appear arbitrary or subjective, thus undermining the rigor initially promised by the mathematical approach. In contrast, the taxonomy of centaur models, based on the distinctions between monotonic vs. non-monotonic systems and open vs. closed systems, offers a straightforward and effective approach. These categories are formally straightforward yet conceptually effective and can be empirically validated, providing solid guidelines for system design.

In its current form, Intellectology seems to grapple with a fundamental issue common to theories premised on the concept of artificial general intelligence: the notion that intelligence can be fully addressed through purely computational means. This perspective has long been a mental trap for various AI schools since the discipline's inception in the 1950s. As argued in the article *Intelligence May Not Be Computable* [37], general intelligence (as opposed to specialized intelligence as applicable to specific tasks like chess playing, image generation, and financial planning) is deeply complex, inherently socio-cognitive, and collaborative across different agents—natural or artificial. This recognition is a central motivation behind the development of centaur models.

4.3. Multi-Agent Systems

The concept of multi-agent systems (MAS) has a long-standing tradition in AI. Generally, it refers to computer systems composed of multiple interacting intelligent agents that collaborate to solve problems or perform tasks more effectively than a single agent could. These systems utilize various methodologies, including logical inference [38], game theory [39], bio-inspired algorithmic search [40], and machine learning [41]. Traditionally, these approaches have been applied to software agents or homogeneous hardware devices, as seen in swarm robotics experiments. However, no inherent limitation prevents diverse agent types from being integrated into multi-agent scenarios, ranging from simple passive agents like sensors connected via Internet-of-Things (IoT) communication channels to advanced active agents. Humans could participate too, leading to hybrid systems that combine human and synthetic capabilities. This possibility has become increasingly viable with the advent of AI-based agents possessing human-like communication abilities, such as chatbots based on large language models, and the potential for human involvement to maintain adequate control over these evolving scenarios.

The hybrid interaction between human and artificial intelligence in multi-agent systems with humans-in-the-loop shares similarities with, yet remains distinct from, centauric models. In centauric models, human and synthetic intelligence integration results in a composite entity with functional interdependencies, operating as a single cohesive agent. In contrast, multi-agent systems maintain the functional independence of each agent—whether human or synthetic—even within collaborative contexts. This allows the same agent to participate in multiple collaborative environments with different agents, depending on the context.

There are, however, cases where the boundaries between multi-agent systems and centaur models become blurred. The article [42], which seeks to systematize multi-agent systems with hybrid participants (both human and artificial), illustrates both the standard situation of non-overlap and borderline cases through two examples. The first involves

agents such as swarm robots with exploration capabilities in unknown territory, coordinated by a satellite that orients their paths based on local data from the robots. This satellite uses algorithmic methods suggested by a large language model (LLM), which, in turn, collaborates and exchanges information flows with a team of human operators for decision-making purposes. This scenario clearly represents a multi-agent system distinct from centaur models, as evidenced by the agents' broad operational and cognitive independence. The second case, involving human agents providing training feedback to a large action model (LAM) that incorporates and replicates their operational tasks, straddles the line between the two categories. This system can be seen both as a multi-agent system—where the relationship between the human and artificial agents is collaborative for training purposes—and as a centaur model, where human capabilities are enhanced through artificial agents.

In conclusion, adequately relating frameworks for modeling centauric systems and multi-agent systems with hybrid agents holds significant promise. Identifying possible intersections and overlaps between these fields could provide valuable insights for designing hybrid systems, representing an important and promising direction for future research.

5. Conclusions

This paper has explored centaurian models as a robust framework for hybrid intelligence, emphasizing their adaptability and practical application across various domains. By leveraging Herbert Simon's cognitive architecture and systemic perspectives, we have laid the groundwork for understanding how centaurian systems can enhance decision-making, creativity, and other complex tasks by seamlessly integrating human and artificial components. Concrete and effective design guidelines are provided based on the monotonic/non-monotonic and open/closed distinctions.

In comparing centaurian models with the Theory of Extended Mind, Intellectology, and multi-agent systems, we observe that while these frameworks offer valuable insights, centaurian models stand out for their pragmatic approach to integrating human and artificial intelligence. The Theory of Extended Mind, which philosophically extends cognition beyond the brain, may provide a cognitive counterpart for centaurian models. However, given their focus on functional integration—a more general and flexible concept than mind extension—centaurian models are also compatible with the opposing Theory of the Embedded Mind, leaving the choice of alignment open to empirical and methodological validation. Similarly, while Intellectology embarks on a broad, often speculative exploration of cognitive diversity that faces formal and empirical challenges, centaurian models exploit simple yet effective taxonomies to guide the design of hybrid intelligence.

In contrast to multi-agent systems, where agents maintain operational independence, centaurian models emphasize creating a unified entity where humans and artificial intelligence work together as a cohesive whole. The two frameworks represent orthogonal and potentially complementary approaches to hybrid intelligence, one aiming at tighter, organism-like integration and the other facilitating social interaction among agents.

Future research directions should focus on several key areas:

- **Refining Centaur Models:** Further research should explore the refinement of centaur models across various domains, ensuring they remain adaptable and effective as new challenges and opportunities arise.
- **Addressing Ethical and Practical Challenges:** Continued emphasis should be placed on addressing centaur systems' ethical, technical, and organizational challenges. This includes developing methodologies for assessing and mitigating risks, ensuring transparency, and maintaining human oversight.
- **Strengthening Applicability:** Strengthening the applicability of centaur models across real-world domains, particularly in fields involving complex decision-making and creativity, is essential. Empirical studies and practical implementations should accompany this expansion to validate the effectiveness of centaur systems in diverse contexts. While this paper has provided concrete case studies, such as the monotonic cognitive

trading system and the non-monotonic Creative Loop for Art and Design, a future paper will fully develop a comprehensive design methodology and structured workflow, addressing the critical questions centaur systems raise. This future contribution will build on the foundational concepts introduced here, applying them to specific case studies in greater depth and detail, thereby complementing the present work. While the current paper characterizes centaur systems and demonstrates their applicability across various domains, the forthcoming research will provide further methodological insights and practical guidance for designing and implementing centaur systems, enhancing their real-world impact.

By advancing these research directions, centaurian systems can play a crucial role in shaping the future of human-computer interaction, offering innovative solutions that enhance both human and artificial intelligence.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Engelbart, D.C. *Augmenting Human Intellect: A Conceptual Framework*; Report No. AFOSR-3223; Stanford Research Institute: Menlo Park, CA, USA, 1962.
2. Licklider, J.C.R. Man-computer symbiosis. *IRE Trans. Hum. Factors Electron.* **1960**, *1*, 4–11. [CrossRef]
3. Te'eni, D.; Yahav, I.; Zagalsky, A.; Schwartz, D.; Silverman, G.; Cohen, D.; Mann, Y.; Lewinsky, D. Reciprocal Human-Machine Learning: A Theory and an Instantiation for the Case of Message Classification. *Manag. Sci.* **2023**. [CrossRef]
4. Simon, H.A. *The Sciences of the Artificial*, 3rd ed.; MIT Press: Cambridge, MA, USA, 1996. [CrossRef]
5. Liskov, B.H.; Wing, J.M. A behavioral notion of subtyping. *ACM Trans. Program. Lang. Syst.* **1994**, *16*, 1811–1841. [CrossRef]
6. Pareschi, R.; Zappone, F., Chapter 5: Integrating heuristics and learning in a computational architecture for cognitive trading. In *Artificial Intelligence and Financial Behaviour*; Edward Elgar Publishing: Cheltenham, UK, 2023; pp. 111–135. [CrossRef]
7. Viale, R.; Mousavi, S.; Filotto, U.; Alemanni, B. *Artificial Intelligence and Financial Behaviour*; Edward Elgar Publishing: Cheltenham, UK, 2023. [CrossRef]
8. Lo, A.W.; Mamaysky, H.; Wang, J. *Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation*; Working Paper 7613; National Bureau of Economic Research: Cambridge, MA, USA, 2000. [CrossRef]
9. Nadkarni, J.; Ferreira Neves, R. Combining NeuroEvolution and Principal Component Analysis to trade in the financial markets. *Expert Syst. Appl.* **2018**, *103*, 184–195. [CrossRef]
10. Kahn, L.H.; Savas, O.; Morrison, A.; Shaffer, K.A.; Zapata, L. Modelling Hybrid Human-Artificial Intelligence Cooperation: A Call Center Customer Service Case Study. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 3072–3075. [CrossRef]
11. Hummel, K.A.; Pollak, M.; Krahofer, J. A Distributed Architecture for Human-Drone Teaming: Timing Challenges and Interaction Opportunities. *Sensors* **2019**, *19*, 1379. [CrossRef]
12. Freire, S.K.; Panicker, S.S.; Ruiz-Arenas, S.; Rusák, Z.; Niforatos, E. A Cognitive Assistant for Operators: AI-Powered Knowledge Sharing on Complex Systems. *IEEE Pervasive Comput.* **2023**, *22*, 50–58. [CrossRef]
13. Aldorasi, E.; Pareschi, R.; Salzano, F. Creatichain: From Creation to Market. In Proceedings of the Image Analysis and Processing, ICIAP 2023 Workshops—ICIAP International Workshops, Udine, Italy, 11–15 September 2023; Lecture Notes in Computer Science; Foresti, G., Fusiello, A., Hancock, E., Eds.; Springer: Cham, Switzerland, 2024; Volume 14366, pp. 51–62. [CrossRef]
14. Castellano, G.; Fanelli, N.; Scaringi, R.; Vessio, G. Exploring the Synergy Between Vision-Language Pretraining and ChatGPT for Artwork Captioning: A Preliminary Study. In Proceedings of the Image Analysis and Processing—ICIAP 2023 Workshops, Udine, Italy, 11–15 September 2023; Foresti, G.L., Fusiello, A., Hancock, E., Eds.; Springer: Cham, Switzerland, 2024; pp. 309–321.

15. Zicari, R.V.; Brusseau, J.; Blomberg, S.N.; Christensen, H.C.; Coffee, M.; Ganapini, M.B.; Gerke, S.; Gilbert, T.K.; Hickman, E.; Hildt, E.; et al. On Assessing Trustworthy AI in Healthcare. Machine Learning as a Supportive Tool to Recognize Cardiac Arrest in Emergency Calls. *Front. Hum. Dyn.* **2021**, *3*, 673104. [CrossRef]
16. Zicari, R.V.; Ahmed, S.; Amann, J.; Braun, S.A.; Brodersen, J.; Bruneault, F.; Brusseau, J.; Campano, E.; Coffee, M.; Dengel, A.; et al. Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier. *Front. Hum. Dyn.* **2021**, *3*, 688152. [CrossRef]
17. Grasso, M.A.; Pareschi, R. Editorial: Human and artificial collaboration for medical best practices. *Front. Hum. Dyn.* **2022**, *4*, 1056997. [CrossRef]
18. Zicari, R.; Brodersen, J.; Brusseau, J.; Düdder, B.; Eichhorn, T.; Ivanov, T.; Kararigas, G.; Kringen, P.; McCullough, M.; Moslein, F.; et al. Z-Inspection®: A Process to Assess Trustworthy AI. *IEEE Trans. Technol. Soc.* **2021**, *2*, 83–97. [CrossRef]
19. Bex, F.J. AI, Law and beyond. A transdisciplinary ecosystem for the future of AI & Law. *Artif. Intell. Law* **2024**. [CrossRef]
20. Schweitzer, S.; Conrads, M. The digital transformation of jurisprudence: An evaluation of ChatGPT-4's applicability to solve cases in business law. *Artif. Intell. Law* **2024**. [CrossRef]
21. Hewitt, C. Offices Are Open Systems. *ACM Trans. Inf. Syst.* **1986**, *4*, 271–287. [CrossRef]
22. Salazar, J. The brain in light of Edgar Morin's paradigm of complexity. *Complexus* **2021**, *8*, 23–32.
23. Pareschi, R. *Centaur Art: The Future of Art in the Age of Generative AI*, 1st ed.; Springer: Cham, Switzerland, 2024. [CrossRef]
24. Clark, A.; Chalmers, D. The extended mind. *Analysis* **1998**, *58*, 7–19. [CrossRef]
25. Thelen, E.; Smith, L.B. *A Dynamic Systems Approach to the Development of Cognition and Action*; MIT Press: Cambridge, MA, USA, 1994.
26. Ballard, D.H. Animate vision. *Artif. Intell.* **1991**, *48*, 57–86. [CrossRef]
27. Brooks, R.A. Intelligence without representation. *Artif. Intell.* **1991**, *47*, 139–159. [CrossRef]
28. Clark, A. *Being There: Putting Brain, Body, and World Together Again*; MIT Press: Cambridge, MA, USA, 1997.
29. Gombrich, E. *The Image and the Eye: Further Studies in the Psychology of Pictorial Representation*; Collected essays; Phaidon: Oxford, UK, 1982.
30. Adams, F.; Aizawa, K. Defending the bounds of cognition. In *The Extended Mind*; Menary, R., Ed.; MIT Press: Cambridge, MA, USA, 2010; pp. 67–80.
31. Rupert, R.D. Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind? In *The Extended Mind*; Menary, R., Ed.; MIT Press: Cambridge, MA, USA, 2010; pp. 325–353.
32. Farina, M.; Lavazza, A. Mind embedded or extended: Transhumanist and posthumanist reflections in support of the extended mind thesis. *Synthese* **2022**, *200*, 507. [CrossRef]
33. Kurzweil, R. *The Singularity is Near: When Humans Transcend Biology*; Viking Penguin: New York, NY, USA, 2005.
34. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014.
35. Yampolskiy, R.V. The Space of Possible Mind Designs. In Proceedings of the AGI Conference, Cham, Switzerland, 22–25 July 2015.
36. Yampolskiy, R.V. Taxonomies of Intelligence: A Comprehensive Guide to the Universe of Minds. *Seeds Sci.* **2023**. [CrossRef]
37. Denning, P.; Lewis, T. Intelligence May Not Be Computable. *Am. Sci.* **2019**, *107*, 346. [CrossRef]
38. Calegari, R.; Ciatto, G.; Mascardi, V.; Omicini, A. Logic-based technologies for multi-agent systems: A systematic literature review. *Auton. Agents Multi-Agent Syst.* **2021**, *35*, 1. [CrossRef]
39. Rosenschein, J.S.; Zlotkin, G. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*; MIT Press: Cambridge, MA, USA, 1994.
40. Dorigo, M.; Maniezzo, V.; Colorni, A. Ant system: Optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **1996**, *26*, 29–41. [CrossRef]
41. Albrecht, S.V.; Christianos, F.; Schäfer, L. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*; MIT Press: Cambridge, MA, USA, 2024.
42. Borghoff, U.M.; Bottoni, P.; Pareschi, R. A System-Theoretical Multi-Agent Approach to Human-Computer Interaction. In Proceedings of the Computer Aided Systems Theory—EUROCAST 2024—19th International Conference, Las Palmas de Gran Canaria, Spain, 25 February–1 March 2024; Revised Selected Papers; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024.