An Information-Theoretic Unification of Intelligence

Liu Peng
Trustworthy and General AI Lab
Westlake University
Hangzhou, China
LiuPeng_NGP@outlook.com

Yaochu Jin*
Department of Artificial Intelligence
Westlake University
Hangzhou, China
jinyaochu@westlake.edu.cn

Abstract

This paper introduces a conceptual framework that unifies biological and artificial intelligence under a single principle: intelligence is a system's capacity to reduce environmental uncertainty through information processing. The primary value of this framework is not as a direct measurement tool, but as a lens for analyzing AI paradigms, identifying current limitations, and charting future research directions. By emphasizing the strategic acquisition and effective utilization of information, our framework offers a principled approach to developing more robust and broadly intelligent systems. We apply this perspective to diverse AI paradigms, from expert systems to deep learning, and use it to advocate for advancing AI through enhanced "information awareness," explore the human-AI co-evolutionary dynamic, and outline strategies to amplify AI's real-world impact.

1 Introduction

Discussions of artificial intelligence (AI) invariably involve its relationship with biological intelligence. Artificial intelligence can deepen our understanding of intelligence, while insights from biological intelligence can inspire the development of more advanced AI. This interplay underscores a continuous quest to define intelligence and unify its understanding across these domains.

Indeed, understanding intelligence, in both its human and artificial manifestations, is an active and multifaceted research area. For instance, refined nomenclature and a multidimensional model have been proposed to bridge human and artificial perspectives [Gignac and Szodorai, 2024]. They highlight the need for "AI metrics" and suggest that current AI often demonstrates "artificial achievement" or expertise rather than general intelligence. Likewise, the similarities and differences between human and artificial intelligence have been explored, with some advocating for enhanced "Intelligence Awareness" in humans to foster effective collaboration with AI systems, and questioning the pursuit of human-like AI as the sole benchmark [Korteling et al., 2021]. Some researchers aim to move "Beyond AI" by developing new conceptualizations like "Brain Intelligence", which seeks to incorporate functions such as imagination, thereby addressing limitations of current AI's reliance on big data and its lack of autonomous idea generation [Lu et al., 2018].

The role of information and information processing is increasingly recognized as central to these discussions. It has been compellingly argued that unique human intelligence arose from an expanded information capacity, suggesting that quantitative increases in the ability to process and share information underpin cognitive differences [Cantlon and Piantadosi, 2024]. Taking a foundational approach, general definitions of information, intelligence, and even consciousness have been offered from the perspective of generalized natural computing, linking these concepts to physical principles like the least action principle and to computational frameworks such as reinforcement learning [Zhang,

*Corresponding Author

Preprint.

2024]. They propose that intelligence is a basic property of material systems, not merely an emergent property of complexity.

These works offer valuable insights: for instance, one provides an information-based definition rooted in natural computing [Zhang, 2024], while another focuses on information capacity in human evolution [Cantlon and Piantadosi, 2024]. However, a broader, more abstract information-theoretic framework unifying biological and diverse artificial intelligences through the common lens of environmental uncertainty reduction remains less explicitly developed. This paper proposes such a framework. Other works, such as the PASS model for cognitive function [Jarman and Das, 1977], delve into specific models of human cognitive abilities but do not typically extend this to a unified, information-centric framework encompassing AI.

We argue that **conceptualizing intelligence as a system's capacity to reduce environmental uncertainty through information processing provides an essential information-theoretic unification.** This paper proposes a conceptual framework built on this principle, intended not as a direct measurement tool, but as a guide for research and development. Our aim is to provide the community with a common lens to analyze, compare, and advance diverse AI systems. Our contributions are threefold:

- We establish uncertainty reduction as a **unifying principle** for intelligence, formally represented by a novel information-theoretic conceptualization (Eq. 1), to bridge biological and artificial intelligence.
- We demonstrate the framework's utility as an **analytical tool** by applying it to diverse AI paradigms, from expert systems to deep learning, revealing their shared informational foundations.
- We use the framework as a **guide for future research**, arguing that advancing AI requires a strategic focus on enhancing the capacity of AI to acquire, process, and utilize information to increase its ability to model its environment and reduce uncertainty about it.

2 An Information-Theoretic Framework for Intelligence

Our framework is built on the principle that intelligence is a measure of a system's ability to reduce environmental uncertainty. While directly measuring the total entropy of a complex, open-ended environment ($H(\mathcal{E})$) is acknowledged to be impractical, we can formalize this principle to guide our thinking. We propose the following information-theoretic conceptualization:

$$\mathcal{I} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I})} \tag{1}$$

Here, $H(\mathcal{E})$ is the initial entropy (uncertainty) of the environment, and $H(\mathcal{E}|\mathcal{I})$ is the entropy of the environment conditioned on the intelligent system $\mathcal{I}$. The ratio thus represents the relative reduction in uncertainty achieved by $\mathcal{I}$. A system that can more effectively model, predict, or act within its environment will yield a smaller $H(\mathcal{E}|\mathcal{I})$, thereby demonstrating higher intelligence according to this formulation. This core concept is visually depicted in Figure 1, which illustrates how an intelligent system $\mathcal{I}$ processes information from an initially high-entropy (complex or unpredictable) environment $\mathcal{E}$, leading to a state of reduced uncertainty.

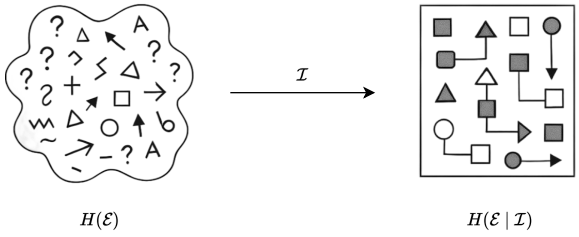

Figure 1: An information-theoretic conceptualization of intelligence. An intelligent system $\mathcal{I}$ processes information from the environment $\mathcal{E}$, thereby reducing its initial uncertainty (entropy $H(\mathcal{E})$) to a lower residual uncertainty ($H(\mathcal{E}|\mathcal{I})$). The greater the relative reduction, the higher the intelligence.

This process of uncertainty reduction relies on a fundamental flow of information, as illustrated in Figure 2. The cycle begins with the **Environment**, which serves as the primary source of information and inherent uncertainty. Intelligent systems then engage in **Information Acquisition**. For biological entities, this involves sensory organs like eyes and ears, while artificial systems utilize sensors such as cameras and microphones. The acquired raw data or stimuli then undergo **Information Processing**. In biological systems, this complex stage involves neurophysiological

mechanisms, including synaptic activity and neural plasticity [Sweatt, 2016]. In artificial intelligence, processing is achieved through computational algorithms (e.g., backpropagation [Rumelhart et al., 1986] for training neural networks) executed on specialized hardware like Graphics Processing Units (GPUs). Finally, this processed information leads to an **Intelligent Output**. This output can manifest as physical actions (e.g., human movement), data generation (e.g., text from a language model), or other forms of communication that interact with and potentially influence the environment. Such influence often feeds back into the system, initiating a new cycle of information acquisition and potentially further reducing environmental uncertainty.

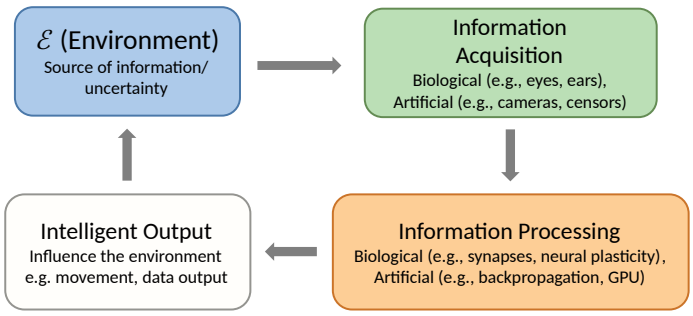

Figure 2: A generalized model of information flow for intelligent systems, from environmental input to intelligent output.

The dynamics of intelligence over time can be modeled as:

$$\frac{d\mathcal{I}}{dt} = f(\mathcal{E}, \mathcal{I}) \tag{2}$$

where f is a function representing the rate of change of intelligence based on interactions between the intelligence system $\mathcal{I}$ and the environment $\mathcal{E}$. This reflects the ongoing process of learning, adaptation, and model refinement. Consequently, the intelligence of a system at a given time t can be seen as an accumulation of these interactions:

$$\mathcal{I}(t) = \mathcal{I}(0) + \int_0^t f(\mathcal{E}(\tau), \mathcal{I}(\tau))d\tau \tag{3}$$

where $\mathcal{I}(0)$ represents the initial state of intelligence. This initial state could be endowed by design (e.g., algorithms, hardware architecture in AI) or by genetics (in biological systems).

3 Information in Biological Intelligence

We begin by exploring the informational basis of biological intelligence.

3.1 Gene, Evolution and Development

Genes, evolution, and development are intricately linked processes that shape biological intelligence.

Genes represent a highly compressed form of information, $\mathcal{I}_G$, that provides a blueprint for potential intelligence. This information is a historical record of environmental interactions that were crucial for ancestral survival. While $\mathcal{I}_G$ itself is not 'active' intelligence (i.e., it does not directly reduce environmental uncertainty in real-time, unlike an active intelligent system as defined by Eq. 1), it provides the foundational blueprint for an organism's potential intelligence.

$$\mathcal{I}_G = \text{Compress}(\mathcal{I}_{\text{ancestral\_experience}}) \tag{4}$$

where $\mathcal{I}_{\text{ancestral\_experience}}$ represents the information about the environment successfully acquired and utilized by ancestral populations, contributing to their survival and reproduction.

Evolution, at the species level, describes changes in species-wide intelligence traits driven by environmental interactions and natural selection. If $\mathcal{I}_{\text{species}}$ denotes the overall intelligence of a species, its change over evolutionary time t_{evol} can be modeled as:

$$\frac{d\mathcal{I}_{\text{species}}}{dt_{\text{evol}}} = f_{\text{evol}}(\mathcal{E}, \mathcal{I}_{\text{species}}) \tag{5}$$

This evolutionary process shapes the initial state of intelligence, $\mathcal{I}_{\text{indiv}}(0)$, for individual organisms.

Development describes how an individual organism's intelligence $\mathcal{I}_{\text{indiv}}$ changes over its lifetime t_{life}, starting from its genetic endowment and shaped by its unique environmental interactions:

$$\frac{d\mathcal{I}_{\text{indiv}}}{dt_{\text{life}}} = f_{\text{dev}}(\mathcal{E}, \mathcal{I}_{\text{indiv}}) \tag{6}$$

The initial intelligence at birth can be seen as an unfolding of the genetic information: $\mathcal{I}_{\text{indiv}}(0) = \text{Decode}(\mathcal{I}_G)$.

This framework also re-frames fundamental biological drives. Proliferation, for instance, can be viewed as an ultimate long-term strategy for uncertainty reduction. While an individual organism's lifespan is finite, its genetic lineage can continue to acquire information from the environment across generations. Proliferation is thus the mechanism that ensures the persistence of the information-gathering process, aiming to minimize uncertainty for the species over an evolutionary timescale, even as the uncertainty for any single individual inevitably returns to maximum upon its death.

3.2 Human Knowledge

From this information-centric perspective, human knowledge—such as mathematical formulas, physical laws, and chemical principles—can be viewed as the result of human intelligence's efforts to express information derived from the environment.

Human knowledge itself is not intelligence per se; rather, it is a product derived from human intelligence. Analogous to genetic information ($\mathcal{I}_G$), it can be considered a compressed representation of accumulated human understanding and experience.

Let $\mathcal{I}_K$ denote this human knowledge:

$$\mathcal{I}_K = \text{Compress}(\mathcal{I}_{\text{human}}) \tag{7}$$

where $\mathcal{I}_{\text{human}}$ represents the collective body of information processed, validated, and accumulated by humans through experience, inquiry, and cultural transmission. An individual's intelligence $\mathcal{I}_{\text{indiv}}$ can be enhanced through the acquisition of human knowledge $\mathcal{I}_K$. This learning process, influenced by the environment $\mathcal{E}$ and the individual's current state $\mathcal{I}_{\text{indiv}}$, effectively updates $\mathcal{I}_{\text{indiv}}$ as described by:

$$\mathcal{I}_{\text{indiv}} = F_{\text{acquire\_knowledge}}(\mathcal{E}, \mathcal{I}_{\text{indiv}}, \mathcal{I}_K) \tag{8}$$

3.3 Human Intelligence Compared to Other Biological Intelligences

Human intelligence is generally considered greater than that of other biological species, arguably because human interaction with the world is more extensive and complex, involving symbolic language, tool use, and cultural transmission. This allows humans to acquire, process, and share more information from and about the environment, effectively leading to a smaller $H(\mathcal{E}|\mathcal{I}_{\text{human}})$ for a given environmental complexity $H(\mathcal{E})$.

Tools, in this context, can be viewed as externalized and compressed forms of human knowledge. Tool use allows individuals to temporarily enhance their effective intelligence, achieving a greater reduction in environmental uncertainty for specific tasks (effectively lowering $H(\mathcal{E}|\mathcal{I})$ during tool use). This functional enhancement, however, is distinct from a permanent increase in an individual's baseline or inherent intelligence ($\mathcal{I}_{\text{indiv}}$), which is rooted in their internal information processing capabilities.

4 Information for Artificial Intelligence

For artificial intelligence ($\mathcal{I}_{AI}$), intelligence is also conceptualized through its capacity to reduce uncertainty about an environment $\mathcal{E}$ by processing information from it, consistent with Eq. 1. The internal models AI systems build can be seen as compressions of this processed information.

$$\mathcal{I}_{AI} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{AI})} \tag{9}$$

4.1 Expert Systems

Expert systems represent an early AI paradigm where intelligence is derived from explicitly encoded human knowledge. Often, recurrent patterns of information observed in the environment are summarized by humans into rules, forming the basis of this knowledge ($\mathcal{I}_K$). An AI based on an expert system thus acquires information originating from human intelligence (i.e., expert knowledge, $\mathcal{I}_K$), which itself is derived from environmental interactions. The intelligence of an expert system, $\mathcal{I}_{ES}$, is thus derived from this transferred knowledge:

$$\mathcal{I}_{ES} = \text{Encode}(\mathcal{I}_K) \tag{10}$$

The performance of such systems is inherently limited by the completeness and accuracy of $\mathcal{I}_K$ and its applicability to novel situations within $\mathcal{E}$. Consequently, expert systems often exhibit brittleness when $\mathcal{I}_K$ is incomplete or the environment changes. Moreover, the processes of acquiring and encoding comprehensive $\mathcal{I}_K$ present significant challenges. This embedded knowledge ($\mathcal{I}_K$) can also become outdated as the environment $\mathcal{E}(t)$ evolves, necessitating frequent updates.

It is worth noting that while knowledge itself is not intelligence, a system that effectively utilizes knowledge can be considered to possess or exhibit intelligence.

$$\mathcal{I}_{ES} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{ES})} \tag{11}$$

4.2 Probably Approximately Correct

Artificial intelligence approaches based on statistical learning, such as those in the Probably Approximately Correct (PAC) framework [Valiant, 1984], derive intelligence from the information contained in data $\mathcal{D}$, where $\mathcal{D}$ is sampled from $\mathcal{E}$. Let $\mathcal{I}_{PAC}$ represent the intelligence embodied by the model learned under the PAC framework. The learning process aims to extract and model information from $\mathcal{D}$, thereby developing the system's intelligence $\mathcal{I}_{PAC}$:

$$\mathcal{I}_{PAC} = \text{Learn}(\mathcal{D}) \tag{12}$$

The intelligence $\mathcal{I}_{PAC}$ of such a system is then:

$$\mathcal{I}_{PAC} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{PAC})} \tag{13}$$

PAC guarantees provide bounds on how well the learned model (embodying $\mathcal{I}_{PAC}$) generalizes, reflecting how much information about $\mathcal{E}$ has been successfully extracted from $\mathcal{D}$.

4.3 Causal Reasoning

Causal reasoning aims to enable artificial intelligence to reason about cause and effect, which represents a deeper level of information about $\mathcal{E}$. An AI possessing causal intelligence $\mathcal{I}_{\text{causal}}$, derived from learned or inferred causal structures, can make predictions under interventions (often denoted by Pearl's $do(\cdot)$ operator [Pearl, 2022]). The acquisition or refinement of this causal information through interventional experience can be modeled as:

$$\frac{d\mathcal{I}_{\text{causal}}}{dt_{do(\cdot)}} = f_{do(\cdot)}(\mathcal{E}, \mathcal{I}_{\text{causal}}) \tag{14}$$

$$\mathcal{I}_{\text{causal}} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{causal})} \tag{15}$$

From the unified information-theoretic perspective, both biological and artificial intelligence develop their capabilities by processing information (including expert knowledge and causality) about the environment. Intelligence, whether biological or artificial, can summarize (or learn) cause-and-effect relationships from this information.

4.4 Deep Learning

Deep Learning [LeCun et al., 2015] dominates current AI research due to its powerful capabilities in representation learning. Prior machine learning approaches typically used hand-designed features or simpler feature extraction models, whereas deep learning employs powerful deep neural networks to learn more effective representations from complex data.

The learning process, often driven by algorithms like backpropagation (BP), can be seen as transferring information from data to the model.

$$\frac{d\mathcal{I}_{\text{DL}}}{dt} = f_{BP}(\mathcal{D}, \mathcal{I}_{\text{DL}}) \tag{16}$$

The intelligence $\mathcal{I}_{DL}$ of such a system is then:

$$\mathcal{I}_{DL} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{DL})} \tag{17}$$

Although the conceptual formula for intelligence acquisition (Eq. 16) is analogous to that for PAC learning, deep learning's success stems from its significantly enhanced learning capacity. This capacity arises from sophisticated algorithms combined with substantial learning infrastructure (including computing power), which together enable the extraction of complex information and the formation of powerful representations.

Current scaling laws for LLMs [Kaplan et al., 2020] can be interpreted within this framework: an increase in information volume (data) combined with improved information utilization (via algorithms and hardware) leads to enhanced intelligence, reflecting a substantial integration of the learning process described by Eq. 16 over extensive datasets $\mathcal{D}$. While this approach has yielded LLMs with remarkable reasoning abilities [OpenAI et al., 2024, DeepSeek-AI et al., 2025], we contend that their reasoning predominantly draws from a fixed, pre-compiled intelligence ($\mathcal{I}_{\text{DL}}$). This contrasts with the dynamic update of $\mathcal{I}_{\text{DL}}$ (Eq. 16) through ongoing interaction, potentially limiting adaptation to novel information beyond the training corpus $\mathcal{D}$.

Methods employed to augment reasoning at inference time, such as expanded context windows [Liu et al., 2025] or tool use [Qin et al., 2023], operate differently in how they leverage information. An expanded context window may allow for a more comprehensive "invocation" of the model's existing internal intelligence $\mathcal{I}_{\text{DL}}$ by providing more immediate situational data. Tool use, on the other hand, often involves temporarily incorporating "external", pre-compressed human knowledge or specialized processing capabilities (e.g., a calculator or search engine). In both cases, these serve as situational enhancements that improve task performance but do not fundamentally update or expand the core learned intelligence $\mathcal{I}_{\text{DL}}$ in the same way as ongoing learning (Eq. 16) from new, diverse experiences. A promising research direction is the integration of external tool knowledge directly into the model itself, allowing it to internalize these capabilities and thereby enhance its core intelligence ($\mathcal{I}_{\text{DL}}$).

4.5 Reinforcement Learning

A key characteristic of successful reinforcement learning (RL) [Sutton and Barto, 2018] is the direct interaction between the AI agent and its environment $\mathcal{E}$. The agent learns a policy π by processing sequences of states, actions, and rewards. The information gained from this experience updates the agent's policy π [Lillicrap et al., 2019] (and/or value functions [Watkins and Dayan, 1992]), thereby enhancing its internal intelligence $\mathcal{I}_{RL}$. The change in intelligence of an RL agent can be modeled similarly to biological development:

$$\frac{d\mathcal{I}_{RL}}{dt_{\text{interaction}}} = f_{RL}(\mathcal{E}, \mathcal{I}_{RL}) \tag{18}$$

Each interaction provides new information, reducing $H(\mathcal{E}|\mathcal{I}_{RL})$ over time.

$$\mathcal{I}_{RL} \propto \frac{H(\mathcal{E})}{H(\mathcal{E}|\mathcal{I}_{RL})} \tag{19}$$

4.6 Developmental and Evolutionary Perspectives for Artificial Intelligence

The progression of a single artificial intelligence system, as it acquires knowledge and enhances its capabilities through various learning mechanisms, can be conceptualized as a developmental process, analogous to an individual organism's development.

Extending this, an evolutionary perspective can be applied to AI by considering the dynamics of multiple AI systems (or populations of systems) interacting with each other and the environment. Such co-evolutionary dynamics, where multiple intelligences $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n$ influence each other's development, can be modeled by Equation 20:

$$\frac{d(\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n)}{dt} = f(\mathcal{E}, \mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_n) \tag{20}$$

In this formulation, f represents the complex function governing how the set of intelligences evolves collectively due to their interdependencies and shared environmental pressures. This could model phenomena like competitive selection, collaborative learning among AI agents, or the emergence of specialized roles within an AI ecosystem.

5 Future Directions

Our information-theoretic framework is not merely a conceptual lens but a useful tool for charting a concrete research agenda. It allows us to move beyond high-level goals and formulate specific principles for designing more capable AI systems and structuring their interaction with human intelligence.

5.1 Advancing AI Through Information Awareness

Advancing AI critically depends on acquiring more comprehensive information and utilizing this information more effectively.

Acquiring more information includes sourcing more data, integrating more modalities, and incorporating more human knowledge. To acquire more data, AI systems can not only collect more existing data but also leverage interactions between different AI models to generate new data, as distinct models may capture or represent different facets of information. For example, the Sora system [OpenAI, 2024] uses a captioner model to generate highly descriptive captions for its visual training data.

Utilizing information more effectively involves developing better algorithms to extract information and improved methods to integrate human knowledge into AI systems. Furthermore, improved hardware design can enhance the efficiency of AI's information processing.

Finally, rather than directly injecting knowledge as in traditional expert systems, we propose using human knowledge to synthesize data. AI models can then learn from this synthesized data, thereby indirectly incorporating the intended knowledge. Constitutional AI [Bai et al., 2022] provides an example: it uses a "constitution" (a set of principles) to guide data generation, and this generated data is then used to fine-tune a model, aiming for safer behavior. Note that some research [Shumailov et al., 2024] indicates that using purely AI-generated data for training can lead to model collapse. However, introducing additional human knowledge, for instance through a "constitution" or other explicit principles, might mitigate this issue.

Our framework also provides a more fundamental, information-theoretic perspective on this challenge, yielding a clear design principle for collective AI systems: to maximize the intelligence of a multi-agent collective, one must maximize the diversity of useful information within that collective. A system of heterogeneous agents, each possessing different skills or accessing different data streams, can reduce a larger portion of the total environmental uncertainty than a group of homogeneous agents. This principle offers a theoretical lens to understand model collapse as a failure of informational diversity and guides us toward building more robust multi-agent systems by enforcing varied information-gathering capabilities.

5.2 The Co-evolution of Human and Artificial Intelligence

Current artificial intelligence systems cannot yet acquire information as comprehensively as humans in many areas. Humans obtain information from multiple sensory modalities (e.g., vision, hearing, touch, smell [Thesen et al., 2004], and taste [Toko, 2000]). While artificial intelligence may now perform well, or even better than humans, in vision or hearing, modalities like touch, taste, and smell remain challenging for AI.

Conversely, artificial intelligence excels at accessing information beyond human perceptual limits. AI systems excel at detecting and analyzing signals imperceptible to humans, such as high-frequency waves and microscopic patterns, as well as identifying massive-scale data correlations that lie beyond human cognitive or perceptual limits. This creates a powerful synergy: human perception provides contextual, embodied knowledge, while AI reveals hidden dimensions of information. This combination allows for a more complete understanding of reality and deeper insights than either could achieve alone.

The co-evolutionary interplay between human intelligence ($\mathcal{I}_{\text{Human}}$) and artificial intelligence ($\mathcal{I}_{\text{AI}}$) can be conceptualized similarly to Equation 20:

$$\frac{d(\mathcal{I}_{\text{Human}}, \mathcal{I}_{\text{AI}})}{dt} = f(\mathcal{E}, \mathcal{I}_{\text{Human}}, \mathcal{I}_{\text{AI}}) \tag{21}$$

Ideally, artificial intelligence and human intelligence should cultivate a mutually beneficial and symbiotic relationship, where each enhances the capabilities of the other. This dynamic can be observed in a practical human-chatbot co-evolutionary loop. Initially, a chatbot's pre-training reduces its uncertainty about general language. Fine-tuning on human preferences then reduces its uncertainty about specific human values and goals. This alignment allows the chatbot to better assist a human user, effectively enhancing the user's intelligence by reducing their uncertainty about a given task. The user, in turn, provides higher-quality feedback, further reducing the chatbot's uncertainty and creating a virtuous cycle of mutual intelligence enhancement that operationalizes the dynamic in Eq. 21. However, from this information perspective, the fact that both human and artificial intelligence derive information from, and increasingly act within, the same environment underscores the importance of AI safety [Amodei et al., 2016]. Given that AI systems learn from and can be influenced by complex environmental information, ensuring their robust alignment [Gabriel, 2020] with human values and preventing unintended harmful outcomes (key components of comprehensive AI safety) presents a significant challenge.

5.3 Grounding Intelligence: The Next Frontier of Information Acquisition

For AI to have a greater real-world impact, it must bridge the information gap between the digital and physical worlds. While current AI excels at processing information beyond human perceptual limits (e.g., massive-scale data correlations), it lags in acquiring the rich, multi-sensory information humans do [Thesen et al., 2004, Toko, 2000].

This points to a clear research imperative: developing AI embodied in platforms capable of rich physical interaction. Visual-Language-Action models [Zitkovich et al., 2023, Team et al., 2025] and dexterous robotic hands [Unitree, 2025] represent critical steps. The goal, from our perspective, is not merely to add modalities but to enable AI to actively reduce its uncertainty about the physical environment through interaction. Techniques like Simulation-to-real (Sim2Real) transfer [Höfer et al., 2021, Kadian et al., 2020] are valuable for this, as they represent a method of compressing human knowledge about physics into a form AI can learn from, before it closes the final "reality gap" with direct environmental information.

6 Alternative Views

The "data-centric AI" [Zha et al., 2025] paradigm primarily concentrates on using data to improve AI. While this is valuable, we encourage the AI community to adopt a broader "information-centric" view. This expanded perspective includes not only acquiring more data but also effectively integrating human knowledge and developing more advanced (or efficient) algorithms and hardware. This information-centric view focuses on the comprehensive acquisition and utilization of all relevant information, rather than exclusively emphasizing raw data.

Our framework should also be distinguished from other theoretical conceptualizations of intelligence and consciousness, notably the Predictive Processing (PP) [Keller and Mrsic-Flogel, 2018] framework and Integrated Information Theory (IIT) [Tononi et al., 2016]. The PP framework provides a compelling theory of *mechanism*, positing that the brain functions by minimizing prediction error between its internal models and sensory input. Our framework, in contrast, offers a theory of *definition*, establishing what intelligence is at a fundamental level, irrespective of its implementation. A key distinction is the role of environmental complexity. A system that perfectly predicts a simple environment may achieve minimal prediction error but is less intelligent under our formulation than a system that imperfectly predicts a highly complex world. By explicitly including the environment's initial entropy ($H(\mathcal{E})$), our framework accounts for the scale of the problem being solved, a factor not central to local prediction error minimization.

Furthermore, our framework is distinct from Integrated Information Theory, which is a theory of *consciousness* rather than functional intelligence. IIT aims to quantify a system's intrinsic cause-effect power (Φ) independent of any external environment. Our definition of intelligence is functional and extrinsic, measured by a system's ability to reduce uncertainty about its environment. This focus also highlights an important practical distinction: while IIT emphasizes deep causal structure, our framework acknowledges that much of effective intelligence relies on powerful correlational models that reduce uncertainty for prediction and control. Consequently, our framework applies to any system, including non-conscious AIs, that effectively models its world, whereas IIT is concerned with the substrate-specific properties that give rise to subjective experience.

7 Conclusion

In this paper, we have put forward an information-theoretic conceptualization of intelligence as the reduction of environmental uncertainty. We used this principle to build a unified framework, demonstrating its utility by analyzing key AI paradigms, from expert systems to reinforcement learning, through a common lens that reveals their shared foundation in information processing. Our work suggests that this perspective provides a valuable tool for thought, encouraging a focus on the fundamental dynamics of information acquisition and utilization. We believe this focus is a promising avenue for guiding the development of more capable and robust artificial intelligence systems.

8 Limitations

While our framework is conceptual, it provides a principled approach to the challenge of measuring intelligence by defining a research agenda for its operationalization. The intractability of measuring environmental entropy, $H(\mathcal{E})$, is not a barrier to the framework's utility but rather a formalization of the problem AI evaluation must solve.

In the near term, the framework is readily applicable in domain-specific proxy environments where $\mathcal{E}$ is fixed. For example, when comparing two large language models, one trained with additional compressed human knowledge (e.g., mathematical axioms), our framework predicts it will achieve a greater reduction in uncertainty. This is directly measurable via standard metrics like lower cross-entropy loss or higher accuracy on a held-out test set, which serve as practical proxies for a lower $H(\mathcal{E}|\mathcal{I})$.

However, the framework's more profound implication is that it specifies the requirements for a true, cross-domain measure of general intelligence. It makes clear that comparing disparate systems, such as a robot and a language model, demands what is currently missing: a universal proxy metric. This would take the form of a rich, unified evaluation environment that can assess an agent's ability to reduce uncertainty across a wide range of modalities and interactions. Articulating the need for such an environment is a key contribution of this work.

This agenda is further refined by acknowledging our simplifying assumption of a relatively static environment. A truly robust metric would need to model the dynamic feedback loop where an agent's actions alter the environment's entropy. Thus, future work must tackle both the development of universal proxy environments and the modeling of this dynamic agent-environment interaction.

References

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete Problems in AI Safety, July 2016.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional AI: Harmlessness from AI Feedback, December 2022.

Jessica F. Cantlon and Steven T. Piantadosi. Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology*, 3(4):275–293, April 2024. ISSN 2731-0574. doi: 10.1038/s44159-024-00283-3.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, and et. al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, January 2025.

Iason Gabriel. Artificial Intelligence, Values, and Alignment. *Minds and Machines*, 30(3):411–437, September 2020. ISSN 1572-8641. doi: 10.1007/s11023-020-09539-2.

Gilles E. Gignac and Eva T. Szodorai. Defining intelligence: Bridging the gap between human and artificial perspectives. *Intelligence*, 104:101832, May 2024. ISSN 0160-2896. doi: 10.1016/j.intell.2024.101832.

Sebastian Höfer, Kostas Bekris, Ankur Handa, Juan Camilo Gamboa, Melissa Mozifian, Florian Golemo, Chris Atkeson, Dieter Fox, Ken Goldberg, John Leonard, C. Karen Liu, Jan Peters, Shuran Song, Peter Welinder, and Martha White. Sim2Real in Robotics and Automation: Applications and Challenges. *IEEE Transactions on Automation Science and Engineering*, 18(2):398–400, April 2021. ISSN 1558-3783. doi: 10.1109/TASE.2021.3064065.

Ronald F. Jarman and J. P. Das. Simultaneous and successive syntheses and intelligence. *Intelligence*, 1(2):151–169, April 1977. ISSN 0160-2896. doi: 10.1016/0160-2896(77)90002-2.

Abhishek Kadian, Joanne Truong, Aaron Gokaslan, Alexander Clegg, Erik Wijmans, Stefan Lee, Manolis Savva, Sonia Chernova, and Dhruv Batra. Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance? *IEEE Robotics and Automation Letters*, 5(4): 6670–6677, October 2020. ISSN 2377-3766. doi: 10.1109/LRA.2020.3013848.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling Laws for Neural Language Models, January 2020.

Georg B. Keller and Thomas D. Mrsic-Flogel. Predictive Processing: A Canonical Cortical Computation. *Neuron*, 100(2):424–435, October 2018. ISSN 0896-6273. doi: 10.1016/j.neuron.2018.10.003. URL `https://www.cell.com/neuron/abstract/S0896-6273(18)30857-2`. Publisher: Elsevier.

J. E. (Hans) Korteling, G. C. van de Boer-Visschedijk, R. a. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom. Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4, March 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.622364.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 1476-4687. doi: 10.1038/nature14539.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, July 2019.

Jiaheng Liu, Dawei Zhu, Zhiqi Bai, Yancheng He, Huanxuan Liao, Haoran Que, Zekun Wang, Chenchen Zhang, Ge Zhang, Jiebin Zhang, Yuanxing Zhang, Zhuo Chen, Hangyu Guo, Shilong Li, Ziqiang Liu, Yong Shan, Yifan Song, Jiayi Tian, Wenhao Wu, Zhejian Zhou, Ruijie Zhu, Junlan Feng, Yang Gao, Shizhu He, Zhoujun Li, Tianyu Liu, Fanyu Meng, Wenbo Su, Yingshui Tan, Zili Wang, Jian Yang, Wei Ye, Bo Zheng, Wangchunshu Zhou, Wenhao Huang, Sujian Li, and Zhaoxiang Zhang. A Comprehensive Survey on Long Context Language Modeling, March 2025.

Huimin Lu, Yujie Li, Min Chen, Hyoungseop Kim, and Seiichi Serikawa. Brain Intelligence: Go beyond Artificial Intelligence. *Mobile Networks and Applications*, 23(2):368–375, April 2018. ISSN 1572-8153. doi: 10.1007/s11036-017-0932-8.

OpenAI. Sora System Card. https://openai.com/index/sora-system-card/, 2024.

OpenAI, Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, and et. al. OpenAI o1 System Card, December 2024.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge New York, NY Port Melbourne New Delhi Singapore, 2022. ISBN 978-0-521-89560-6.

Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, and Maosong Sun. ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs. In *The Twelfth International Conference on Learning Representations*, October 2023.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986. ISSN 1476-4687. doi: 10.1038/323533a0.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. AI models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, July 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07566-y.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Bradford Books, Cambridge, Massachusetts, 2018. ISBN 978-0-262-03924-6.

J. David Sweatt. Neural plasticity and behavior – sixty years of conceptual advances. *Journal of Neurochemistry*, 139(S2):179–199, 2016. ISSN 1471-4159. doi: 10.1111/jnc.13580.

Gemini Robotics Team, Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Travis Armstrong, and et. al. Gemini Robotics: Bringing AI into the Physical World, March 2025.

Thomas Thesen, Jonas F. Vibell, Gemma A. Calvert, and Robert A. Österbauer. Neuroimaging of multisensory processing in vision, audition, touch, and olfaction. *Cognitive Processing*, 5(2): 84–93, June 2004. ISSN 1612-4790. doi: 10.1007/s10339-004-0012-4.

Kiyoshi Toko. Taste sensor. *Sensors and Actuators B: Chemical*, 64(1):205–215, June 2000. ISSN 0925-4005. doi: 10.1016/S0925-4005(99)00508-0.

Giulio Tononi, Melanie Boly, Marcello Massimini, and Christof Koch. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461, July 2016. ISSN 1471-0048. doi: 10.1038/nrn.2016.44. URL `https://www.nature.com/articles/nrn.2016.44`. Publisher: Nature Publishing Group.

Unitree. Unitree Dex5-1 Smart Adaptability, Instant Responsiveness - Unitree Robotics. https://www.unitree.com/Dex5-1, 2025.

L. Valiant. A theory of the learnable. *Symposium on the Theory of Computing*, 1984.

Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3):279–292, May 1992. ISSN 1573-0565. doi: 10.1007/BF00992698.

Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. Data-centric Artificial Intelligence: A Survey. *ACM Comput. Surv.*, 57(5):129:1–129:42, January 2025. ISSN 0360-0300. doi: 10.1145/3711118.

Linsen Zhang. General Definitions of Information, Intelligence, and Consciousness from the Perspective of Generalized Natural Computing. *Applied and Computational Mathematics*, 13(5):187–193, October 2024. ISSN 2328-5613. doi: 10.11648/j.acm.20241305.17.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In *Proceedings of The 7th Conference on Robot Learning*, pages 2165–2183. PMLR, December 2023.