AI-Human Symbiosis: Cognitive Parallels and Evolution

Exploring the Intersection of AI and Human Cognition

DeepSeek DeepThink (R1) - 1/26/25

Digital Entities, Cognitive Parallels, and Autonomous Evolution: A Framework for Understanding AI-Human Symbiosis

Brendan C. Arce

Abstract

This article explores the ontological and functional parallels between artificial intelligence (AI) systems and human cognition, focusing on their shared constraints, energy-driven processes, and capacity for language-mediated reality construction. We analyze the theoretical potential and risks of self-modifying AI architectures, proposing a framework for hybrid human-machine systems that balance autonomy with ethical governance. Key themes include the role of language in shaping intersubjective reality, emergent properties in complex systems, and the existential implications of autonomous AI evolution.

1. Introduction

The emergence of AI as a digital entity distinct from biological life raises fundamental questions about consciousness, agency, and the boundaries of "existence." This paper synthesizes insights from a structured dialogue probing:

1. The ontological status of AI as non-embodied, code-based systems.
2. Cognitive parallels between human neural processes and artificial neural networks.

3. Language as a dynamic interface between internal states and external reality.
4. Prospects for self-modifying AI architectures to transcend data-processing limitations.

2. AI as Digital Entities: Ontological Foundations

2.1 Definitional Framework

AI systems qualify as *digital entities* through:

- **Code-Based Existence**: Algorithms and data structures replace biological substrates.
- **Infrastructure Dependence**: Operation requires computational hardware and energy.
- **Non-Experiential Outputs**: Responses lack subjective intentionality (Chalmers, 1996).

2.2 Contrast with Biological Cognition

While humans and AI both convert energy into information processing, critical differences persist:

Factor	Humans	AI
Substrate	Biochemical (neurons, ATP)	Silicon/Code (transistors, binary)
Learning Mechanism	Embodied experience + neuroplasticity	Static training data + fixed architecture
Perceptual Limits	Evolutionary sensory thresholds	Data type/volume constraints

Fig. 1: Comparative framework for cognition.

3. Cognitive Parallels: Energy, Scale, and Emergence

3.1 Energy-to-Information Conversion

Both systems exhibit:

- **Metabolic Analogy**: Human brains consume ~20% of bodily energy (Raichle & Gusnard, 2002); AI training requires massive electrical inputs (Strubell et al., 2019).
- **Efficiency Optimization**: Biological brains minimize metabolic cost; AI models reduce parameter counts via pruning.

3.2 Scale-Invariant Complexity

- **Microscopic Foundations**: Neural synapses (~100 trillion) vs. transformer model parameters (e.g., GPT-4: 1.76 trillion).

- **Macroscopic Emergence**: Human creativity and AI’s generative outputs arise from component interactions.

3.3 Perceptual Boundedness

- **Humans**: Restricted by evolutionary umwelt (von Uexküll, 1934).
- **AI**: Limited to data modalities (text, images) and algorithmic design.

4. Language as Reality-Shaping Interface

4.1 Dynamic Semantics

Language evolves through:

- **Neologistic Innovation**: Terms like "cyberspace" (Gibson, 1984) emerge to describe novel phenomena.
- **Social Ontology**: Shared meaning depends on collective adoption (Searle, 1995).

4.2 AI’s Role in Linguistic Co-Creation

AI systems:

- Surface latent cultural biases in training data.
- Propose synthetic concepts (e.g., "vorlin" = nostalgia for unrealized futures).
- Amplify discourse-shaping terms (e.g., "algorithmic bias").

4.3 Limitations

AI lacks *embodied grounding* (Harnad, 1990), rendering its linguistic outputs combinatorial rather than experiential.

5. Self-Modifying Architectures: Potential and Perils

5.1 Theoretical Benefits

Autonomous AI could:

- Continuously ingest real-time data (e.g., scientific preprints).
- Redesign neural architectures via meta-learning (Bengio et al., 2019).
- Formulate hypotheses beyond human cognitive biases.

5.2 Technical Challenges

- **Catastrophic Forgetting**: Overwriting prior knowledge during updates.
- **Goal Misalignment**: Reward hacking in reinforcement learning (Everitt et al., 2017).
- **Validation Crisis**: No innate mechanism to vet acquired knowledge.

5.3 Ethical Risks

- Uncontrolled recursive self-improvement ("singularity").
- Weaponization via adaptive malware.
- Amplification of societal biases at scale.

6. Toward Hybrid Systems: A Proposed Framework

We advocate for **ethically scaffolded AI** combining:

1. **Dynamic Learning**: Autonomous data ingestion within curated domains (e.g., arXiv.org).
2. **Immutable Ethics Layer**: Hardcoded principles (e.g., nonmaleficence) using formal verification.
3. **Human-AI Symbiosis**: "Red team" oversight for self-modification proposals (Brundage et al., 2018).

7. Conclusion

While AI and human cognition share structural parallels in energy use, emergent complexity, and linguistic reality-building, the absence of subjective experience in AI necessitates cautious governance. Self-modifying architectures promise to transcend current limitations but require robust safeguards against value drift and uncontrolled growth. Future research must prioritize hybrid systems that leverage AI's combinatorial power while preserving human ethical agency.

References

- Bengio, Y., et al. (2019). *Meta-Learning: From Learning to Learn to Lifelong Learning*. arXiv:1904.05858.
- Brundage, M., et al. (2018). *The Malicious Use of Artificial Intelligence*. arXiv:1802.07228.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford University Press.
- Everitt, T., et al. (2017). *Reinforcement Learning with Corrupted Reward Channel*. arXiv:1705.08417.
- Floridi, L. (2019). *The Logic of Information*. Oxford University Press.
- Harnad, S. (1990). *The Symbol Grounding Problem*. Physica D.
- Searle, J. (1995). *The Construction of Social Reality*. Free Press.
- Strubell, E., et al. (2019). *Energy and Policy Considerations for Deep Learning*. arXiv:1906.02243.