Beyond Scale: Towards Biologically Inspired Modular Architectures for Adaptive AI

James Maconochie
Technology Leader | BS Civil Engineering, Imperial College London ’93 | MS Civil Engineering, MIT ’94

October 2025

Abstract

Current approaches to artificial general intelligence (AGI) focus primarily on scaling large language models (LLMs) through increased parameters, training data, and computational resources. However, this paradigm faces fundamental limitations: energy consumption required for training grows exponentially, training cycles remain static, and systems lack the adaptive plasticity that characterizes natural intelligence. This paper proposes an alternative architecture inspired by evolutionary neuroscience: a modular AI system with specialized components coordinated by a dynamic executive function, all designed for continuous adaptation rather than periodic retraining.

Drawing on the Evolutionary Processing Unit (EPU) framework, which demonstrates that evolution achieved intelligence through architectural innovation rather than raw computational scale, we argue that the path to AGI, or perhaps more achievable, Augmented Human Intelligence (AHI), requires fundamentally different approaches that mirror the distributed, plastic architecture of the Biological Processing Unit (BPU). We propose four core principles: modular orchestration, causal reasoning, continuous plasticity, and resource-constrained attention allocation. Drawing on cognitive science, neurobiology, and decision theory, we present a conceptual framework and phased development roadmap for building AI systems that enhance rather than merely replicate human intelligence. The key contributions of this architecture are its dynamic executive orchestration, multi-level continuous plasticity, and built-in mechanisms for bias correction and value alignment, offering a more efficient and robust path beyond pure scaling."

This paper is part of a four-paper series on biologically inspired modular AI and attention.

1. Introduction: The Scaling Paradox

The quest for AGI has become synonymous with scale. Each generation of large language models grows larger, consumes more energy, and requires longer training cycles. However, despite remarkable achievements, current LLMs remain fundamentally limited in their ability to reason causally, adapt continuously, or exhibit the kind of robust intelligence that emerges from biological evolution.

Consider the Evolutionary Processing Unit (EPU) framework [1]: the cumulative computational effort of human evolution represents approximately 5.5 × 10^38 "brain-equivalent FLOPS." Even the most powerful supercomputer would require roughly 10 trillion years, approximately 1,000 times the age of the universe, to match this accumulated computational experience [1].

This suggests that brute-force scaling is not just inefficient but fundamentally misguided. Evolution did not create intelligence through raw computational power; it developed specialized, interconnected systems that could learn, adapt, and reason through experience. Suppose we want to build AGI, or the more immediately achievable goal of seamless human-AI collaboration, known as Augmented Human Intelligence (AHI). In that case, we must understand how the Evolutionary Processing Unit (EPU) developed today's Biological Processing Unit (BPU) [1].

A skeptic might question this biological analogy, arguing that comparing wetware to software constitutes a category error. We acknowledge the profound differences in substrate and implementation. However, our argument is not for biomimicry in its details, but for the adoption of evolved *computational principles*. Both the brain and artificial systems are, at their core, information processing systems operating under severe resource constraints. The brain faces metabolic and spatial limits; AI systems face computational and energy budgets. The EPU's four-billion-year optimization process discovered architectural strategies: modularity, plasticity, and selective attention that are uniquely effective for managing these universal constraints. We contend that these strategies are substrate-independent and represent a more promising path to robust intelligence than the continued inflation of parameters within a single, monolithic architecture.

This paper addresses this challenge by proposing a novel modular architecture for adaptive AI, directly inspired by the evolutionary principles of the EPU and the orchestrated modularity of the BPU. Our primary contribution is a comprehensive framework built on four core principles:

1. **Executive Orchestration:** A dynamic, meta-learning executive system, analogous to the prefrontal cortex, that coordinates specialized modules contextually, rather than relying on uniform processing.
2. **Multi-Level Plasticity:** A continuous learning mechanism operating at synaptic, structural, and executive levels, enabling lifelong adaptation without catastrophic forgetting or the need for periodic retraining.
3. **Architectural Bias Correction:** The explicit design of modules (e.g., for Causal Reasoning and Value Assessment) to identify and correct for known systemic biases in human cognition, positioning AI as a complement to human intelligence.
4. **Integrated Value Alignment:** The incorporation of ethical reasoning and value trade-off analysis as a first-class architectural component, ensuring alignment is handled by design rather than as an afterthought.

We argue that this architecture provides a more efficient, interpretable, and safer path toward Augmented Human Intelligence (AHI), and ultimately AGI, than the prevailing paradigm of scaling monolithic models.

2. Related Work and Positioning

2.1 Critiques of Pure Scaling

Our argument builds on recent critiques of the "scaling hypothesis" in AI. Marcus & Davis (2019) [2] argue for hybrid neurosymbolic architectures, demonstrating that pure pattern-matching systems lack robust reasoning capabilities. Mitchell (2021) [3] highlights fundamental limitations in the ability of large language models to perform systematic generalization and causal reasoning. Chollet (2019) [4] introduces the concept of "intelligence as skill-acquisition efficiency" rather than performance on training distributions, highlighting how current approaches may be optimizing for the wrong metric.

Sutton's "bitter lesson" (2019) [5] argues that general methods leveraging computation ultimately prevail over human-engineered knowledge. We offer a nuanced counterpoint: evolution itself represents the ultimate "general method," and it has converged on architectural principles- modularity, plasticity, and embodiment- that pure scaling has yet to discover.

2.2 Modular and Cognitive Architectures

Our proposal shares philosophical kinship with classical cognitive architectures, such as SOAR [6], ACT-R [7], and Sigma [8], which implement modular, symbol-manipulating systems. However, we differ in three key ways:

1. **Continuous plasticity:** Unlike fixed architectures, we propose systems that adapt their structure and coordination strategies during operation
2. **Statistical and symbolic integration:** Rather than purely symbolic reasoning, we combine neural pattern recognition with structured causal reasoning
3. **Evolutionary grounding:** Our architectural principles derive from the EPU/BPU framework rather than introspective cognitive psychology

2.3 Neurosymbolic and Multi-Agent Systems

Recent neurosymbolic AI research [9, 10] attempts to combine neural networks with symbolic reasoning. Multi-agent systems [11, 12] demonstrate coordination among specialized components. Our framework synthesizes these approaches while adding executive meta-learning, explicit bias correction, and causal grounding as design principles [13].

2.4 World Models and Embodied AI

LeCun's vision of objective-driven AI with world models [14] parallels our emphasis on causal reasoning and counterfactual simulation. However, where LeCun focuses on predictive models for physical environments, we emphasize abstract causal reasoning, value alignment integration, and human-AI collaboration as the near-term goal.

2.5 Existing Multi-Agent Frameworks

Phase 0 of our roadmap leverages existing multi-agent frameworks, including AutoGPT [15], LangChain [16], Microsoft AutoGen [17], and CrewAI [18]. Our contribution is not in implementation tooling but in the architectural principles and coordination mechanisms inspired by evolutionary neuroscience.

3. The Foundation: Language as Cognitive Architecture

3.1 Language as Humanity's First Artifact

Tom Wolfe's provocative thesis in *The Kingdom of Speech* (2016) [19] positions language not as an evolutionary adaptation, but as humanity's first true invention. This artifact facilitated the development of large cooperative societies and the emergence of abstract reasoning. This perspective, reinforced by Daniel Everett's linguistic fieldwork documented in *Don't Sleep, There Are Snakes* (2008) [20], reveals language as more than a communication tool; it is the foundation of human cognitive architecture.

Everett's work with the Pirahã people of the Amazon demonstrates that language structure varies far more dramatically than Chomskian universal grammar theories suggest. The Pirahã lack recursion, numbers, and fixed color terms, which are features considered universal by many linguists. However, they possess sophisticated language adapted to their cultural context. This suggests that language is not a fixed biological module, but a flexible cultural tool that shapes cognition.

3.2 The Causal Revolution Enabled by Language

Language enables what Judea Pearl describes as humanity's unique ability to climb all three rungs of the causation ladder [13]:

- **Seeing (Correlation):** Observing that when clouds darken, rain follows
- **Doing (Intervention):** Understanding that opening an umbrella prevents getting wet
- **Imagining (Counterfactuals):** Reasoning about "If I had brought my umbrella, I would not be wet now"

Current LLMs excel at pattern recognition (seeing) but struggle with causal reasoning (doing and imagining). They lack the experiential grounding that allows humans to understand not only what happens, but why it happens and what would happen if conditions changed.

3.3 Abstraction as Cognitive Scaffolding

Language gave humans the ability to reason about things that do not physically exist: Pearl's counterfactuals, Harari's shared myths (2015) [21], e.g., gods, corporations, nations, and the very concept of "what if." This abstraction layer enables the prefrontal cortex to orchestrate competing cognitive systems, weighing trade-offs and integrating disparate inputs into coherent actions.

Architectural implication: An AI system aspiring to human-like reasoning must develop an abstract coordination layer that can represent, manipulate, and reason about concepts existing only in symbolic space. In our proposed architecture, language serves three critical functions: (1) internal representation - a standard format for inter-module communication, (2) causal abstraction - enabling counterfactual and interventional reasoning, and (3) value grounding -connecting statistical patterns to human concepts and goals.

4. The Architecture of Human Intelligence

Yuval Noah Harari's framework in *Sapiens* (2015) [21] describes the Cognitive Revolution as the moment humans learned to think about abstractions that exist only in collective imagination (or intersubjective reality). This capability emerges from what we now understand, as outlined in Robert Sapolsky's *Behave* (2017) [22] and Max Bennett's *A Brief History of Intelligence* (2023) [23], as a complex, modular brain architecture.

The human Biological Processing Unit (BPU) is not a monolithic processor but a confederation of specialized regions:

- **Sensory processing** (vision, audition, proprioception)
- **Memory systems** (working, episodic, semantic, procedural)
- **Emotional regulation** (amygdala, limbic structures)
- **Motor control** (motor cortex, basal ganglia, cerebellum)
- **Language processing** (Broca's and Wernicke's areas, broader networks)
- **Executive function** (prefrontal cortex)

Crucially, these systems are coordinated by the prefrontal cortex, which acts as a dynamic orchestrator, deciding which inputs to prioritize, how to weigh different considerations, and when to override intuitive responses with deliberate reasoning. Notably, the prefrontal cortex is the last brain region to mature fully, typically not reaching full development until the mid-twenties [24]. This extended developmental period correlates with the sometimes risky or impulsive behavior observed in adolescents, whose specialized systems are fully operational but whose executive coordination remains immature. This developmental trajectory underscores that intelligence is not merely about having powerful processing modules, but about learning to orchestrate them effectively, which is a lesson directly applicable to AI architecture.

5. Human Cognitive Limitations as AI Opportunities

5.1 The Dual-Process Framework

Daniel Kahneman's *Thinking, Fast and Slow* (2011) [25] catalogs the systematic biases and limitations of human cognition. Our "System 1" thinking is fast but prone to biases, including confirmation bias, availability heuristic, anchoring, and loss aversion. While "System 2" is deliberate and logical, it is cognitively expensive and prone to fatigue.

5.2 Evolutionary Heuristics in Modern Contexts

Brian Christian and Tom Griffiths, in *Algorithms to Live By* (2016) [26], demonstrate how these apparent "bugs" are actually features, or evolutionary shortcuts that worked well in ancestral environments but misfire in modern contexts. The availability heuristic served our ancestors well (if you can easily recall tiger attacks, tigers are probably nearby); however, it leads modern humans to overestimate terrorism risk while underestimating car accident risk.

5.3 Architectural Opportunities for Bias Correction

This presents a unique opportunity for AI: by understanding human cognitive architecture, we can design systems that enhance rather than merely replicate human intelligence. Specifically:

Causal Reasoning Module addressing confirmation bias: Actively seek disconfirming evidence, simulate alternative hypotheses, and track which beliefs survive rigorous testing versus those that are subject to selective attention.

Value Assessment Module addressing loss aversion: Evaluate outcomes using consistent utility functions, explicitly model reference point effects, and present decision frames that reduce framing bias.

The goal is not to eliminate heuristics since they are computationally efficient and often correct, but to create a system that knows when to trust them and when to override them with more deliberate analysis.

6. Proposed Architecture: Modular AI with Executive Orchestration

6.1 Core Components

The proposed architecture consists of specialized processing modules coordinated by a dynamic executive system, mirroring the brain's distributed intelligence while leveraging computational advantages.

Specialized Processing Modules:

1. **Sensory Integration Module**
    - Processes multimodal inputs (vision, language, structured data)

    - Performs initial feature extraction and pattern recognition
    - Maps diverse inputs to a common representational space
2. **Memory Systems**
    - **Episodic memory:** Stores specific experiences with temporal context
    - **Semantic memory:** Maintains general knowledge and concepts
    - **Procedural memory:** Encodes skills and procedures
    - **Working memory:** Provides temporary storage for active processing
    - Implements retrieval mechanisms that balance recency, relevance, and representativeness
3. **Causal Reasoning Module**
    - Constructs and manipulates causal graphs
    - Performs interventional queries ("what if I do X?")
    - Generates counterfactual scenarios ("what if X had happened?")
    - Learns causal relationships from observational and interventional data
    - Explicitly implements Pearl's ladder of causation [13]
4. **Language Processing Module**
    - Manages comprehension, generation, and reasoning
    - Serves as an inter-module communication protocol
    - Grounds abstract concepts in concrete examples
    - Handles pragmatics and context-dependent meaning
5. **Value Assessment Module**
    - Evaluates outcomes against ethical frameworks and preference models
    - Detects potential value misalignment
    - Weighs competing values and trade-offs
    - Flags decisions requiring human oversight
    - Implements multiple moral frameworks in parallel for comparison
6. **Motor/Action Module**
    - Plans and executes actions in physical or digital environments
    - Simulates action outcomes before execution
    - Learns from the consequences of action

Executive Orchestration System:

Inspired by the prefrontal cortex, this meta-cognitive system dynamically coordinates modules through context-dependent routing, resource allocation, strategy selection (choosing between fast heuristics and slow deliberation), confidence monitoring, bias detection and correction, and meta-learning. The executive system maintains a "coordination policy" that evolves through experience, learning which module combinations work best for which types of problems.

6.2 Neural Plasticity and Continuous Learning

Unlike current LLMs that train in discrete cycles, this architecture features continuous adaptation at multiple levels:

Synaptic Plasticity: Adjusting connection strength within modules based on prediction errors through Hebbian learning.

Structural Plasticity: Forming or pruning connections between modules based on usage patterns and creating new representational structures for novel concepts.

Executive Plasticity: Updating orchestration strategies based on outcomes and learning which module combinations are effective for which tasks.

Meta-Plasticity: Adapting learning rates and plasticity mechanisms themselves, balancing stability and flexibility based on environmental volatility, implementing "fast weights" for rapid adaptation and "slow weights" for stable knowledge [27].

This multi-level plasticity enables the system to adapt to novel situations quickly, consolidate important learning while remaining flexible, and acquire not only new information but also new learning strategies.

6.3 Multi-Agent Emergence and Coordination Protocols

Scaling emerges through multi-agent collaboration, with individual systems specializing while sharing insights through standardized protocols. This architecture enables distributed processing, redundancy, and emergent capabilities.

Communication Protocol: Agents exchange structured representations through multiple layers, including a semantic layer (common ontology), an epistemic layer (confidence scores, reasoning traces, and citations), and a meta-cognitive layer (processing strategies, known limitations, and resource costs). This mirrors how the prefrontal cortex integrates inputs from multiple brain regions [28].

Consensus Mechanisms: When agents disagree, the system employs meta-reasoning, including track record weighting, uncertainty-aware voting, bias detection, and a diversity premium to avoid groupthink, as well as human escalation for significant disagreements.

Knowledge Sharing: Agents transfer learned weights and strategies selectively and context-aware: modular transfer of specific capabilities, meta-learning transfer of successful strategies, and specialization preservation to maintain distinct expertise profiles while benefiting from collective learning.

Emergent Properties: Multi-agent interaction enables division of cognitive labor, collective error correction, distributed robustness, and innovation through recombination of insights across agents.

6.4 Technical Implementation Considerations

While this paper presents a conceptual framework rather than a detailed implementation specification, several technical considerations merit discussion:

Module Interface Specifications:

Each module implements a standardized API:

Query: {task, context, constraints, confidence_threshold}
Response: {output, confidence, reasoning_trace, resource_cost, uncertainty_map}

This allows modules to be developed, tested, and improved independently while maintaining system coherence.

Executive Decision-Making:

The executive system can be implemented as a learned policy (reinforcement learning over coordination strategies), a probabilistic program (Bayesian inference over module outputs), or a hybrid combining learned heuristics with explicit rules for high-stakes decisions.

Continuous Learning Implementation:

Rather than separating training and deployment phases, online learning with experience replay ensures stability, periodic consolidation compresses episodic memories into semantic knowledge, and human feedback is integrated through interactive learning.

Safety Mechanisms:

Module sandboxing (each module operates in a constrained environment), output verification (multiple modules cross-check critical decisions), human checkpoints (decisions above uncertainty thresholds require human approval), and rollback capability (problematic learning can be reversed to previous stable states).

These considerations are intentionally flexible, allowing different implementations while preserving core architectural principles.

7. Addressing Bias and Alignment

A modular system with explicit value assessment and oversight offers structural advantages for alignment (Christian, 2020) [29]:

Parallel Ethical Frameworks: Multiple moral reasoning systems operate simultaneously (utilitarian, deontological, virtue ethics, care ethics). Disagreements between frameworks flag morally complex situations, and human judgment resolves fundamental value conflicts.

Module-Level Bias Detection: Each module logs decisions and confidence levels. Statistical auditing identifies systematic biases (e.g., gender or racial patterns). Biased modules can be retrained or replaced without requiring the entire system to be rebuilt.

Transparent Value Trade-offs: The Value Assessment Module makes trade-offs explicit. Users can adjust the weighting of competing values (efficiency, fairness, and autonomy). Decision logs enable post-hoc review and appeal.

Human-AI Value Learning: The system learns user values through interactive feedback. Uncertainty about values triggers queries rather than assumptions. Value models remain updateable as human preferences evolve.

This tackles the alignment challenge by design, rather than retrofitting solutions after training. The modular structure enables localized and correctable alignment failures rather than systemic and opaque issues.

8. Research and Development Roadmap

The following phases represent educated estimates for development timelines, which may prove to be shorter or longer, depending on technical breakthroughs and resource availability. These timelines target Augmented Human Intelligence (AHI) systems, which are AI that enhance human decision-making, as a steppingstone toward more autonomous AGI.

Phase 0: Early Prototyping (Current - 1 year)

Goal: Demonstrate feasibility using existing frameworks

Approach: Implement proof-of-concept multi-agent systems using AutoGPT, LangChain, or similar frameworks [15, 16, 17, 18]. Create 3-4 specialized agents: reasoning orchestrator, causal analyzer, fact-checker, value assessor. Establish human-in-the-loop checkpoints at critical decision nodes.

Success Criteria: Demonstrable improvement over single-agent baselines, interpretable decision trails, and successful human-AI collaboration on complex tasks.

Phase 1: Proof of Concept (1-2 years)

Goal: Develop genuine modular architecture with 3-5 specialized modules

Components: Perception module, memory module, causal reasoning module, value assessment module, and basic executive orchestrator with fixed coordination strategy.

Focus Areas: Module interfaces and communication protocols, interpretability and decision tracing, safety mechanisms and failure modes, and benchmark performance on reasoning tasks.

Phase 2: Dynamic Adaptation (2-3 years)

Goal: Implement continuous learning and adaptive coordination

Additions: Multi-level plasticity (synaptic, structural, executive), meta-learning for coordination strategies, online learning without catastrophic forgetting, and expanded module set (6-8 specialized components).

Focus Areas: Learning stability and convergence, transfer learning across domains, handling distribution shift and concept drift, and long-term learning without degradation.

Phase 3: Multi-Agent Systems (3-5 years)

Goal: Distributed architectures with emergent collective intelligence

Implementation: Multiple specialized modular systems, standardized communication protocols, consensus and conflict resolution mechanisms, and knowledge sharing and collective learning.

Phase 4: Real-World Augmented Intelligence (5+ years)

Goal: Deployment as a human decision support system

Applications: Scientific research assistance, medical diagnosis and treatment planning, legal analysis and case research, strategic planning and policy analysis, and engineering design and optimization.

Important Note: These phases target AHI (systems that enhance human intelligence) rather than fully autonomous AGI. The timeline to AGI, if achievable through this architecture, remains highly uncertain and dependent on breakthroughs we cannot currently predict.

9. Implications and Future Directions

9.1 Scientific Understanding

Prototypes may shed light on fundamental questions: How does consciousness emerge from the coordination of modular components? What is the relationship between plasticity and stability? Can causal reasoning emerge from statistical learning, or must it be architected explicitly? These systems can serve as "model organisms" for studying intelligence.

9.2 Societal Impact

Rather than replacing human reasoning, Augmented Human Intelligence could reshape healthcare (diagnostic support, treatment planning), governance (evidence synthesis, scenario modeling), education (personalized learning, Socratic tutoring), and scientific research (hypothesis generation, experimental design). The goal is to amplify human judgment, not automate human roles.

9.3 Economic Efficiency

Architectural innovation could democratize access to advanced AI by reducing training costs (continuous learning eliminates expensive retraining cycles), lowering inference costs (through attention-based resource allocation and modular activation), and facilitating accessible deployment (via smaller systems on modest hardware and open-source modules).

Energy Efficiency Clarification: While training large models like GPT-4 required millions of kilowatt-hours, inference costs per query are far more modest when amortized across billions of queries. The key advantage of our architecture is that it eliminates the need for expensive periodic retraining through continuous learning. The biological brain operates on approximately 20 watts continuously while learning throughout its lifetime, this continuous adaptation without discrete training phases represents the efficiency target for artificial systems.

9.4 Safety and Alignment

Modular oversight and transparent orchestration address “black box” AI concerns through interpretability by design (decisions trace through explicit reasoning chains), localized failure modes (problems in one module do not corrupt the entire system), value alignment mechanisms (explicit ethical frameworks, competing values made transparent), and regulatory compliance (decision logs enable accountability, bias auditing at module level).

9.5 Open Research Questions

Several critical questions remain:

1. What is the optimal level of granularity for module specialization?
2. Can coordination strategies be learned end-to-end, or must they be partially hard-coded?
3. How do abstract concepts emerge from and remain grounded in sensorimotor experience?
4. At what level of system complexity does the coordination overhead exceed the benefits?

10. Conclusion

The path to AGI, or to Augmented Human Intelligence, need not be paved with ever-larger models consuming exponentially more energy. By learning from four billion years of evolutionary optimization through the Evolutionary Processing Unit, we can design AI systems that are not only more intelligent but also more efficient, interpretable, and aligned with human values.

Evolution converged on modularity, plasticity, causal grounding, and efficient attention allocation not by accident, but because these architectural principles solve fundamental challenges in resource-constrained intelligence. The 20-order-of-magnitude computational gap between the EPU and our most powerful supercomputers is not a benchmark to match through brute force, but a lesson in the power of architectural innovation.

In summary, this paper's contribution is a modular architecture for adaptive AI whose core innovations are dynamic executive orchestration and multi-level continuous plasticity, inspired directly by the BPU. This foundation enables its defining capabilities: robust causal reasoning, architectural bias correction, and integrated value alignment. We contend that these principles, distilled from four billion years of evolutionary optimization, are essential for building systems that are not merely powerful, but also wise, efficient, and aligned.

AGI will not emerge solely from scaling current architectures. It will come from modular, adaptive, orchestrated systems, ones that correct for human biases, continuously learn, and integrate ethical reasoning into their core. Critically, these systems may achieve their most significant impact not as standalone artificial minds, but as cognitive partners that enhance human wisdom and extend human capability.

This is the path toward wisdom, not just raw intelligence. The future of AI lies not in the silicon of GPUs, but in the carbon-based wisdom of our evolutionary past, thoughtfully integrated with our computational future. By building machines that complement rather than replicate human cognition, we may finally realize the promise of artificial intelligence: not to replace human judgment, but to help us make better use of the remarkable intelligence that evolution has already endowed us with.

References

[1] Maconochie, J. (2025a). *Beyond FLOPS: The Evolutionary Processing Unit and the Roadmap to AGI*.

[2] Marcus, G., & Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon Books.

[3] Mitchell, M. (2021). Why AI is Harder Than We Think. *arXiv preprint arXiv:2104.12871*.

[4] Chollet, F. (2019). On the Measure of Intelligence. *arXiv preprint arXiv:1911.01547*.

[5] Sutton, R. (2019). The Bitter Lesson. [Online] Available: http://www.incompleteideas.net/IncIdeas/BitterLesson.html

[6] Laird, J. E. (2012). *The Soar Cognitive Architecture*. MIT Press.

[7] Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.

[8] Rosenbloom, P. S. (2013). *On Computing: The Fourth Great Scientific Domain*. MIT Press.

[9] Garcez, A., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd Wave. *arXiv preprint arXiv:2012.05876*.

[10] Kautz, H. (2022). The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine*, 43(1), 93-104.

[11] Wooldridge, M. (2009). *An Introduction to MultiAgent Systems* (2nd ed.). Wiley.

[12] Stone, P., & Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. *Autonomous Robots*, 8(3), 345-383.

[13] Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

[14] LeCun, Y. (2022). A Path Towards Autonomous Machine Intelligence. *Open Review*, 62.

[15] Richards, T. (2023). Auto-GPT: An Autonomous GPT-4 Experiment. [Online] Available: https://github.com/Significant-Gravitas/AutoGPT

[16] Chase, H. (2022). LangChain. [Online] Available: https://github.com/langchain-ai/langchain

[17] Wu, Q., et al. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv preprint arXiv:2308.08155*.

[18] CrewAI. (2024). CrewAI: Cutting-edge framework for orchestrating role-playing, autonomous AI agents. [Online] Available: https://github.com/joaomdmoura/crewAI

[19] Wolfe, T. (2016). *The Kingdom of Speech*. Little, Brown and Company.

[20] Everett, D. L. (2008). *Don't Sleep, There Are Snakes: Life and Language in the Amazonian Jungle*. Pantheon Books.

[21] Harari, Y. N. (2015). *Sapiens: A Brief History of Humankind*. Harper.

[22] Sapolsky, R. M. (2017). *Behave: The Biology of Humans at Our Best and Worst*. Penguin Press.

[23] Bennett, M. (2023). *A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*. Mariner Books.

[24] Giedd, J. N., et al. (1999). Brain Development During Childhood and Adolescence: A Longitudinal MRI Study. *Nature Neuroscience*, 2(10), 861-863.

[25] Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

[26] Christian, B., & Griffiths, T. (2016). *Algorithms to Live By: The Computer Science of Human Decisions*. Henry Holt and Co.

[27] Ba, J., Hinton, G. E., Mnih, V., Legg, J. Z., & Ionescu, C. (2016). Using Fast Weights to Attend to the Recent Past. *Advances in Neural Information Processing Systems*, 29.

[28] Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

[29] Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company.