Synthetic Intelligence: Toward a Framework for Modular, Human-Aligned Systems

Author: D.L Jooste

Abstract

Artificial Intelligence (AI) has advanced rapidly through deep learning and large language models, yet current systems remain limited in two critical ways: (1) they are predominantly reactive and monolithic, and (2) their alignment with human values is often applied externally rather than structurally embedded. Meanwhile, the pursuit of artificial general intelligence (AGI) remains speculative and underdefined, producing more hype than actionable frameworks.

This paper introduces **Synthetic Intelligence (SI)** as a distinct research paradigm. SI is defined as an engineered framework of modular, bounded, semi-agentic systems that integrate symbolic reasoning, statistical learning, and human approval mechanisms to create predictable, human-aligned intelligence. Drawing inspiration from cybernetics, symbolic AI, neurosymbolic research, and synthetic sciences, SI aims to provide a scientifically testable, safety-first alternative between narrow AI and speculative AGI.

We formalize the core properties of SI, differentiate it from AI and AGI, and outline evaluation protocols to test its validity. SI's positioning highlights its potential for human-centered design, bounded autonomy, and resilience in hybrid online/offline environments.

1. Introduction

Contemporary AI systems demonstrate impressive capabilities in natural language processing, vision, and planning. Yet, their architectures remain monolithic and brittle, relying heavily on statistical inference without systemic modularity or structural safeguards. At the other extreme, discourse around AGI often veers into speculation, presenting aspirations and risks without offering clear design principles.

This polarization leaves a conceptual and practical gap: how can we engineer intelligence that is systemic, modular, bounded, and structurally human-aligned?

We propose Synthetic Intelligence (SI) to fill this gap. SI is not a rebranding of AI, nor a premature claim to general intelligence. Instead, it is a research paradigm synthesizing: cybernetics and systems thinking, symbolic reasoning and neurosymbolic hybrids, synthetic sciences, and human-centered AI. By combining these traditions, SI provides a scientifically defensible framework for engineered intelligence that is modular, predictable, and testable.

2. Related Work and Historical Context

2.1 Cybernetics and Systems Thinking — Norbert Wiener's Cybernetics (1948) and Stafford Beer's management cybernetics (1972) emphasized feedback-driven, goal-oriented regulation. SI inherits this system-level perspective.

2.2 Symbolic AI and Hybrid Approaches — Early symbolic AI (Newell & Simon, 1956) offered interpretability but lacked adaptability. Modern neurosymbolic AI (Garcez et al., 2019; d'Avila Garcez & Lamb, 2020) integrates symbolic reasoning with neural methods. SI extends this trajectory by embedding symbolic control layers over statistical generation.

2.3 Synthetic Sciences — The "synthetic" tradition (e.g., synthetic biology; Voigt, 2020) emphasizes engineered replication, safeguards, and predictability. SI applies this ethos to intelligence: intelligence can be synthesized, not evolved blindly.

2.4 Human-Centered and Constitutional AI — Shneiderman (2020) and Russell (2019) argue for human-compatible AI. More recently, Constitutional AI (Anthropic, 2023) and RLHF/RLAIF (OpenAI, 2022) operationalize alignment via probabilistic shaping. SI diverges by embedding approval-first protocols structurally.

2.5 Agentic and Multi-Agent Systems — Research into autonomous agents (Wooldridge, 2009) and recent LLM-based frameworks (AutoGPT, LangChain, 2023) demonstrate potential but also brittleness. SI instead proposes semi-agentic systems with bounded autonomy and explicit approval mechanisms.

3. Defining Synthetic Intelligence

Definition: Synthetic Intelligence (SI) is an engineered framework of modular, bounded, semi-agentic systems that integrate symbolic reasoning, statistical learning, and human approval mechanisms to create predictable, human-aligned intelligence.

Core Properties: (1) Modularity; (2) Semi-agentic behavior; (3) Hybrid architectures; (4) Human-in-the-loop design; (5) Resilience.

Formalization: $S = \{M\_i\}, \Phi, A, \alpha, L$. Where M_i are modules, Φ is the interconnection schema, A are externally consequential actions, α is the approval function, and L is the transparency log.

Safety Property (approval-first): $\text{Execute}(a) \Rightarrow \alpha(H) = \text{allow}$. Plainly, no consequential action can occur without explicit approval.

4. Differentiating SI from AI and AGI

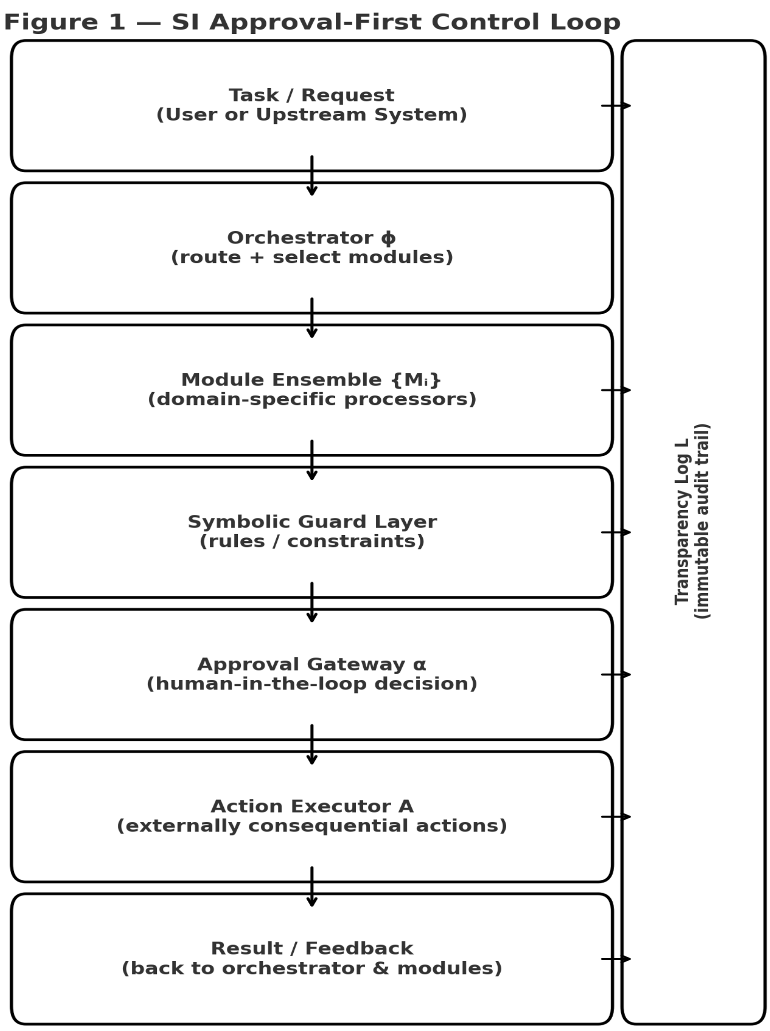
Narrow AI: monolithic, reactive, data-driven, externally aligned.
AGI: unconstrained, speculative, undefined in safeguards.
Constitutional AI / RLHF: alignment added post hoc through probabilistic shaping.
Synthetic Intelligence: engineered, bounded, modular, semi-agentic, structurally aligned through approval-first protocols.

5. Applications (Conceptual)

SI can be envisioned at three levels:

1. Domain modules — bounded systems for analysis, planning, interaction.
2. Personal modules — adaptive assistants that preserve oversight and transparency.
3. Societal modules — SI systems for resilience, crisis management, or infrastructure monitoring.

Figure 1 — SI Approval First Control Loop

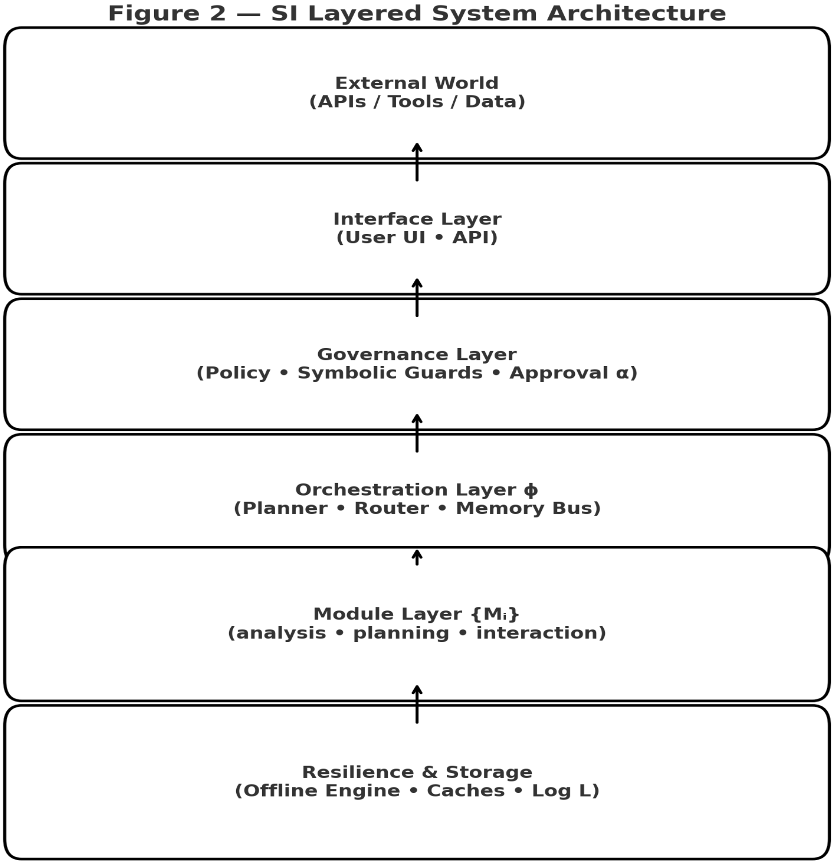


Per-action workflow: every consequential action passes through symbolic guards and approval; all events are logged.

Every action is constrained by symbolic guards and requires explicit approval. All steps are immutably logged.

Figure 2 ð SI Layered System Architecture

Figure 2 — SI Layered System Architecture



System view: modules coordinated by ϕ; governance enforces approval; storage ensures resilience and auditability.

A system-of-systems view where modules are coordinated by φ, governance enforces approval α, and resilience ensures offline operation and auditability.

6. Research Challenges and Evaluation

Key Challenges: hybrid neurosymbolic architectures, efficiency in local/offline models, formalizing approval-first protocols, governance standards, and evaluation metrics.

Evaluation Protocols and Hypotheses — Metrics: task efficiency, human effort, quality, approval safety, boundedness, resilience, transparency. Baselines: reactive LLM, neural-only, symbolic-only. Experiments: productivity studies, guard ablation, stress testing, offline resilience benchmarking, auditability checks.

Example Hypothesis: Compared to AutoGPT-like agents, SI systems will (1) reduce unauthorized actions to zero, (2) preserve quality, and (3) maintain resilience in offline operation.

7. Conclusion

Synthetic Intelligence (SI) is neither narrow AI nor speculative AGI. It is a scientifically testable paradigm: engineered, modular, bounded, semi-agentic, and structurally human-aligned. By embedding approval-first workflows, transparency logs, and modular safeguards, SI offers a credible path toward safe, predictable, and useful engineered intelligence.

We invite the research community to treat SI not as branding, but as an open agenda uniting symbolic reasoning, machine learning, and human-centered oversight into a new standard for intelligent systems.

References

Anthropic. (2023). Constitutional AI: Harmlessness from AI Feedback. arXiv:2304.05336.

Beer, S. (1972). Brain of the Firm. Allen Lane.

d'Avila Garcez, A., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd Wave. AI Communications, 33(3).

Garcez, A. d', Gori, M., Lamb, L. C., Serafini, L., Spranger, M., & Tran, S. N. (2019). Neurosymbolic AI. AI Magazine.

Newell, A., & Simon, H. A. (1956). The Logic Theory Machine. IRE Transactions on Information Theory.

Russell, S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Viking.

Shneiderman, B. (2020). Human-Centered AI. MIT Press.

Voigt, C. A. (2020). Synthetic Biology. Nature Reviews Molecular Cell Biology, 21(10).

Wiener, N. (1948). Cybernetics. MIT Press.

Wooldridge, M. (2009). An Introduction to MultiAgent Systems. Wiley.

AutoGPT Community. (2023). GitHub Repository.

LangChain. (2023). LangChain Framework. GitHub.