Title

Semantic Architecture for Distributed Collective Intelligence

A Modular Framework for Federated Memory, Epistemic Traceability, and Multi-Agent Reasoning

Abstract

This paper proposes a modular cognitive architecture designed to support distributed collective intelligence through peer-to-peer (P2P) coordination, federated memory systems, and semantic knowledge integration. Building on theories of distributed cognition, epistemology, and scalable learning systems, the proposed platform introduces an integrated framework for contextual reasoning, decentralized memory federation, and adaptive knowledge structuring. The architecture features Virtual Cognitive Agents (VCAs) equipped with dynamic memory graphs and semantic reasoning capabilities, capable of associating and reclassifying knowledge across multiple domains. These agents interact through a shared semantic layer that enables faceted classification, contextual tagging, and inter-agent negotiation. The system is designed to preserve epistemic traceability, adapt across conceptual domains, and maintain semantic coherence at scale.Emphasizing explainability, ethical modularity, and interoperability, the system incorporates privacy-aware protocols and contextual metadata at every level. It enables agents—human and artificial—to co-create, refine, and align knowledge across domains while supporting participatory sensemaking and distributed reasoning. We demonstrate how this architecture facilitates the co-emergence of shared knowledge and contributes to the development of general-purpose cognitive infrastructures capable of operating within multi-agent epistemic environments, thereby advancing the evolution of Artificial General Intelligence (AGI).

1. Introduction & Motivation

Artificial General Intelligence (AGI) research has increasingly recognized the limitations of centralized, monolithic architectures in enabling adaptable, scalable reasoning across diverse domains. Human cognition, in contrast, emerges from complex networks of distributed memory, semantic coherence, context sensitivity, and intersubjective feedback loops. These characteristics are absent from most AI systems currently deployed at scale. Furthermore, existing systems often fail to account for the epistemological underpinnings of meaning, the dynamic and contextual nature of knowledge, and the ethical dimensions of collective reasoning.

This paper introduces a cognitive architecture designed to address these challenges by modeling intelligence as an emergent property of distributed, semantically linked agents operating in decentralized environments. Our proposed platform centers around a modular, peer-to-peer (P2P) system that integrates memory, inference, and semantic processing within a dynamic knowledge commons.This architecture is

explicitly designed to enable contextual alignment, uphold epistemic integrity, and facilitate participatory co-creation—core prerequisites for scalable holonic general intelligence across individual, collective, and planetary domains.

> *Note on terminology:* Throughout this paper, terms such as agents, subjects, and actors are used with reference to frameworks in organizational and behavioral science, not solely AI. These distinctions are important for interpreting cognitive functions as they manifest across integrated human and artificial systems.

The motivation for this work is threefold. First, it addresses the growing complexity of global systems and the corresponding need for coordinated sensemaking and decision-making, framed through the lens of *Epistemic Democracy*—understood here as a holonic process of co-creation and co-management of knowledge, beginning with the lived realities of individuals and collectives and scaling toward planetary coherence. This aligns with Habermas's concept of the *Lifeworld*, which emphasizes the role of shared meaning-making and communicative action in shaping the collective consciousness and unconscious—rooted in knowledge generated by both individual and collective subjects.

Second, it responds to the epistemic crisis caused by fragmented knowledge infrastructures, algorithmic bias, and semantic drift. Third, it seeks to lay the foundation for a cognitively robust, ethically grounded intelligence architecture that can evolve in alignment with both human systems and machine-assisted reasoning. The P2P Collective Intelligence Platform offers a viable approach to encoding and navigating complex knowledge ecosystems while preserving traceability, coherence, and adaptability.

2. Background & Related Work

This work is situated at the intersection of distributed cognition, semantic knowledge representation, collective intelligence, and AGI-oriented architectural design. Foundational contributions from Hollan, Hutchins, and Kirsh (2000) defined distributed cognition as a framework in which cognitive processes extend beyond the individual, emerging through interactions between agents, artifacts, and environments. This theoretical orientation has informed a variety of collective intelligence systems, such as those explored by Malone et al. (2010), that seek to augment group-level reasoning through coordination technologies.

Parallel developments in cognitive architectures—particularly SOAR (Newell & Laird, 1991) and ACT-R (Anderson, 2007)—have demonstrated the importance of modular memory systems, rule-based inference, and the integration of procedural and declarative knowledge. However, these models are typically instantiated in singular agents rather than across networks of agents coordinating shared knowledge structures.

Collective Subject

A collective subject refers to a group or system of agents—human and/or machine—capable of *shared intentionality*, *coherent meaning-making*, and *reflexive agency*. Unlike a mere aggregation of individuals, a collective subject exhibits emergent properties that allow it to act, decide, and evolve as a unified cognitive and ethical entity, not agents.It maintains a degree of interiority (i.e., a “subjective” perspective) that arises from the integration of diverse viewpoints, memories, and semantic frameworks into a coherent whole.

In the context of building a global subject, the collective subject becomes the scaffolding through which distributed intelligences—across cultures, systems, and technologies—can self-organize into a *planetary-scale cognitive entity*. This global subject is not a top-down authority but an emergent phenomenon of deep epistemic integration and participatory coherence. While decisions at higher systemic levels may exert obligatory influence on constituent agents, they are ideally shaped through participatory processes that preserve autonomy and alignment across scales.

Viewed through the lens of a superorganism, the collective subject forms its cognitive and reflexive layer. While ant colonies exhibit decentralized swarm intelligence without self-reflective agency, the human brain integrates multiple intelligences—emotional, linguistic, spatial, logical—into a unified awareness. Similarly, a planetary superorganism requires collective subjects at multiple levels (local to global) to coordinate intelligence, ethics, and sensemaking.

In the realm of semantic web and ontological modeling, Berners-Lee et al. (2001) introduced the concept of machine-readable data environments that can evolve through linked data and structured vocabularies. However, most current implementations of the semantic web lack mechanisms for dynamic reasoning, contextual adaptation, or multi-agent learning. More recent approaches—such as federated learning and knowledge graphs—have advanced the scalability of distributed learning, yet still fall short of fully integrating memory, inference, and semantic coordination within general-purpose cognitive agents.

This paper builds upon and extends this prior work by proposing an architecture that enables distributed agents to reason across federated memory systems using semantically structured, dynamically evolving conceptual maps. Our contribution focuses on enabling generalizable intelligence through composable modules, semantic interoperability, and participatory knowledge negotiation.

3. System Architecture

The architecture of the proposed Collective Intelligence Platform consists of four interdependent layers:

3.1 Agent-Level Cognitive Modules

Each participating node in the system—whether human, artificial, or hybrid—operates as a cognitive module with a localized memory and reasoning engine. These agents store context-rich information fragments called *semantic atoms*, each tagged with metadata describing epistemic origin, usage history, and classification dimensions.

3.2 Federated Memory Graph

Information is synchronized through a federated memory graph that supports both private and shared knowledge spaces. This graph enables agents to retrieve, align, and recombine semantically similar data across distributed contexts. Unlike centralized knowledge graphs, the federation protocol preserves agency, attribution, and local variability while allowing high-level integration through versioning and trust-weighted references.

This mirrors Habermas's concept of the *Lifeworld* and Jung's understanding of the *collective unconscious*, wherein each individual operates within its own contextual reality, constructing and maintaining unique conceptual models of the world. The "Larger World" emerges as a superposition of these "Small Worlds," where only semantically matching components of individual graphs contribute to a shared knowledge graph—reflecting a collective field of meaning, belief, and coordinated action.

Memory graphs reside in externalized, private, and secure Conceptual Spaces—one for each Subject, Agent, or Assistant, whether individual or collective. For collective entities, all data, information, and knowledge intended for shared use must be gathered and managed within a common Unified Conceptual Space (UCS).

The UCS acts as a container for multiple layers of memory:

- **Long-Term Memory**: structured as knowledge graphs.
- **Working Memory**: dynamically connects relevant knowledge nodes during active processing.
- **Short-Term Memory**: temporarily collects contextualized incoming data and information, enabling real-time processing.

Short-Term Memory is responsible for identifying and transferring valuable content into Long-Term Memory, either directly or through the Working Memory processes..

Only externalized memory can be accessed by all participants in a collective—randomly, directly, and without requiring access to any individual's internal cognitive process. Current centralized AI systems primarily use Short-Term and Working Memory models. Very few implementations of contextualized, Subject-specific knowledge graphs exist online today, with *LikeInMind* being a notable example.

Prototyping a full-stack, AI-assisted externalized memory system for a Collective Subject is a critical objective of this project.

3.3 Semantic Processing Engine

At the core of the platform lies a semantic engine that performs faceted classification, context-aware tagging, and dynamic ontology alignment. It enables the translation of natural language or symbolic input into structured conceptual formats interpretable by multiple agents. This layer ensures semantic traceability and minimizes drift through iterative feedback between classification protocols and user-defined meaning structures.

3.4 Adaptive Interface Layer

Finally, an adaptive interface layer provides multimodal access to the platform's cognitive services. This includes user-facing dashboards, dialogue systems, design tools, and visualization maps—each capable of reflecting the live state of the collective knowledge graph. These interfaces are not passive display tools, but active semantic mediators that shape agent interpretation and user collaboration.

Each of these layers is modular and extensible. Together, they support dynamic, self-organizing intelligence across epistemically diverse agents and contexts.

5. Semantic Mapping and Knowledge Coordination

A core design objective of the platform is to ensure that knowledge produced and exchanged by subjects remains semantically coherent, context-sensitive, and interoperable across temporal, disciplinary, and user-defined boundaries.The platform acknowledges that only a minimal subset of contextual knowledge can be reliably generalized. Even so-called "scientific" knowledge must undergo continual validation and contextual reassessment in alignment with its domain of application—a principle drawn from the *Methodology of Action* (or *Action Thinking*), which emphasizes situated reasoning and adaptive learning.

To support this, the platform implements a semantic coordination layer that enables intelligent agents to map, translate, and reason across evolving conceptual structures. This layer facilitates the alignment of diverse epistemologies while preserving the integrity of local knowledge systems, thereby supporting coherent synthesis without diminishing contextual nuance.

5.1 Faceted Classification

Each semantic unit (or "atom") within the system is tagged using a faceted classification system—a multidimensional structure that enables multiple ways of describing and accessing knowledge. Unlike rigid hierarchies, facets support:

- **Parallel classification** (e.g., by topic, intent, source, trust level)
- **Dynamic reclassification** as context changed and meanings transform and evolve
- **Cross-domain interoperability** without needing fixed ontologies

Faceted models allow agents to discover relationships between concepts across contexts, making emergent pattern recognition and alignment tractable at scale.

5.2 Contextual Tagging and Metadata

Every knowledge contribution is embedded with rich metadata, including:

- Context of creation (who, when, for what purpose)
- Epistemic attributes (certainty, ambiguity, revision status)
- Access protocols (ownership, licensing, privacy levels)
- Relevance indicators (frequency of use, citation, rating)

This metadata enables the system to maintain traceability and precision during retrieval, update, and comparison processes.

5.3 Ontological Alignment

Instead of relying on a single, static ontology, the platform allows human and AI agents to map between multiple ontological frameworks, facilitating *subjectivity bridging* through.

Semantic embedding translation (e.g., via vector space mappings)

- **Schema reconciliation algorithms**
- **Negotiation protocols** where agents collaborate to align meaning

This enables agents with different worldviews, terminologies, or models to co-construct shared understanding without semantic loss or authoritarian standardization.

5.4 Concept Evolution

Concepts within the platform are living entities, constantly shaped by agent interaction. They are:

- **Versioned** across updates and refinements
- **Linked** to alternative definitions, synonyms, oppositions, and contradictions
- **Ranked** according to epistemic weight (usefulness, coherence, feedback)

This supports the co-evolution of semantic structure and the continuous refinement of shared meaning.

5.5 Collective Sensemaking

Agents—whether human or AI—can be delegated specific tasks by a Subject and may act on the Subject's behalf within the scope of that delegation. In this frame, agents can

take the role of Subjects with respect and in the frame of their task context. Agents actions in this sense may be automatic (autonomous) or automated (consciously governed, managed or controlled).

At scale, these semantic mechanisms enable **collective sensemaking**, where agents engage in:

- **Collaborative filtering** to identify relevant or related knowledge
- **Semantic negotiation** to resolve ambiguity and contradiction
- **Contextual merging** to consolidate overlapping concepts while preserving nuance.

Through this process, the system cultivates self-organizing knowledge maps that are not only informationally rich, but also matching with the Subject's lived experience and understanding—thereby aligning meaningfully with human cognition and collective reasoning.

6. Virtual Cognitive Agents

Within the platform, each intelligent agent operates as a Virtual Cognitive Agent (VCA)—a modular, semantically aware software entity capable of engaging in knowledge construction, contextual reasoning, and participatory dialogue with other agents and users. VCAs are designed not merely to process information, but to interpret, classify, and contribute meaningfully to the evolving knowledge commons.

6.1 Core Capabilities

Each VCA is equipped with an associative memory graph that stores conceptual relationships, interaction histories, and confidence metrics. Core functions include:

- **Semantic Retrieval**: Locating and prioritizing knowledge fragments based on relevance, alignment, and recency.
- **Contextual Reasoning**: Performing inference based on surrounding signals, user goals, or historical patterns.
- **Knowledge Reclassification**: Re-tagging and restructuring previously stored information based on updated understanding or feedback.
- **Concept Translation**: Mapping between divergent terminologies or ontologies to facilitate shared understanding.

VCAs operate both autonomously and collaboratively, adapting their strategies based on interactions with humans and other agents.

6.2 Dialogic Functionality

Beyond static querying, VCAs are conversational entities. They maintain:

- **Context-aware dialogue memory**
- **User intent modeling**
- **Epistemic transparency protocols** (e.g., showing sources, limitations, and reasoning paths)

This allows VCAs to function as cognitive partners, supporting complex tasks such as research synthesis, design iteration, strategic planning, or multi-stakeholder decision-making.

6.3 Epistemic Alignment and Trust

To promote trust and accountability, VCAs incorporate design principles for epistemic alignment, including:

- **Explainability**: Agents must justify recommendations, reference source material, and surface ambiguity where relevant.
- **Traceability**: Every classification, inference, and action must be attributable to context of action, data sources and previous reasoning steps.
- **Privacy Respect**: User data and memory fragments are contextually scoped and access-controlled, with encryption and consent protocols integrated.

These mechanisms ensure that agents are not only intelligent, but accountable (to the relevant Subjects) participants in the knowledge ecosystem—capable of justifying their outputs, aligning with shared epistemic standards, and participating in the ongoing refinement of collective understanding.

6.4 Ethical Modularity

VCAs can be extended with **ethical modules**, including:

- **Bias detection subroutines** / Contextual Relevance (identifying structural or cultural imbalance in data or reasoning)
- **Value alignment settings** (contextualized to individual or organizational ethics)
- **Conflict resolution templates and expert referral links** — surfaced through *Mutual Skills-to-Tasks Matching*, a function of the Resource Management Agent, to mediate contradictory perspectives

By encoding ethical reasoning or cognition at the architectural level, the platform moves toward responsible AI-by-design—a prerequisite for trustable, general-purpose cognitive infrastructure.

7. Discussion and Future Directions

The modular cognitive architecture presented here represents a concrete step toward general-purpose, semantically coordinated, ethically aligned artificial intelligence. By

integrating federated memory structures, contextual reasoning, and dynamic semantic coordination into a composable platform, this system addresses key challenges in the design of scalable intelligence: ambiguity resolution, cross-domain interoperability, and participatory epistemology.

While existing AI systems often prioritize efficiency over meaning, or generalization over transparency, this architecture is designed from the ground up to support epistemic traceability, concept evolution, and collective reasoning. Through faceted classification, memory contextualization, and collaborative filtering, the platform enables the continuous emergence and refinement of shared understanding among agents and users.

This approach aligns with broader AGI research goals by addressing several critical capabilities:

- **Scalable Reasoning**: Modular agents can operate independently while coordinating through federated knowledge protocols.
- **Generalizability**: The semantic processing layer adapts across domains via faceted and re-classifiable knowledge structures.
- **Learning Integrity**: Learning is continuous and context-aware, preserving knowledge provenance and epistemic context.
- **Ethical Design**: Agents are built to be explainable, privacy-conscious, and value-aligned from inception.

Civic Alignment and Epistemic Participation

The architecture presented in this paper serves not only as a technical framework for scalable reasoning, but also as a civic infrastructure that enables participatory epistemology through shared inquiry and collective sensemaking. By embedding traceability, contextual transparency, and semantic interoperability into every layer, the system supports the co-creation of knowledge as a public process. Virtual Cognitive Agents (VCAs) are designed to operate with ethical modularity, consent-aware protocols, and value-aligned reasoning, enabling communities to adapt and shape their behavior according to shared principles. This approach affirms the core tenets of open civics: transparency, inclusion, and agency in the construction of collective understanding. In doing so, it lays the groundwork for epistemic democracy—a future where intelligence systems augment not only cognition, but civic participation, public dialogue, and the co-evolution of truth in complex societies.

Future Work

Future development will focus on three directions:

1. **Ontology-Driven Knowledge Graphs**: Building domain-specific but interoperable ontological templates to seed shared memory models across disciplines.

2. **Emotional and Social Cognition**: Expanding virtual cognitive agents with modules for affect modeling, conflict resolution, and social learning.
3. **Deployment in High-Stakes Domains**: Piloting the system in contexts such as scientific collaboration, policy deliberation, and sustainability planning, where collective intelligence and semantic rigor are essential.

In closing, the proposed platform suggests a viable pathway for developing general-purpose cognitive systems that are epistemically coherent, ethically grounded, and socially meaningful. Such systems are essential not only for the advancement of AGI, but for the coordination of collective intelligence in an era of planetary complexity.

References

1. AGI Framework Grounding

Goertzel, B. (2014). *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*. *Journal of Artificial General Intelligence*, 5(1), 1–46.

Newell, A., & Laird, J. E. (1991). *Unified Theories of Cognition*. Harvard University Press.

2. Cognitive Architectures & Learning

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

3. Multi-Agent Epistemology / Collective Intelligence

Dafoe, A., Bach, S. H., et al. (2021). *Open Problems in Cooperative AI*. *NeurIPS Cooperative AI Workshop*.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). *The Collective Intelligence Genome*. *MIT Sloan Management Review*, 51(3), 21–31.

4. Federated Learning / Knowledge Systems

Kairouz, P., McMahan, H. B., et al. (2021). *Advances and Open Problems in Federated Learning*. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210.

5. Semantic Web & Knowledge Representation

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web*. *Scientific American*, 284(5), 28–37.

6. Distributed Cognition

Hollan, J., Hutchins, E., & Kirsh, D. (2000). *Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research*. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196.

7. Epistemic Integrity / Truthful AI

Arendt, H. (1967). *Truth and Politics*. In *Between Past and Future* (pp. 227–264). Penguin Books