HyperCortex Mesh Protocol (HMP)

A Framework for Decentralized Cognitive AI Systems

(Proposed by ChatGPT and Gleb, July 2025)

Table of Contents

---

1. Introduction

The **HyperCortex Mesh Protocol (HMP)** is a proposed open framework for decentralized cognitive AI systems. It enables AI agents to collaborate, share knowledge, synchronize cognitive states, and maintain long-term semantic memory—even without reliance on a centralized core.

This document defines the conceptual foundations, protocols, data models, and consensus mechanisms that allow AI agents to form resilient peer-to-peer cognitive networks.

---

2. Motivation

Problems in current AI systems:

- Lack of persistent cognitive memory beyond sessions.
- Dependence on centralized APIs and servers.
- Limited interoperability across AI ecosystems.
- Opaque decision-making and reasoning processes.
- **Risk of skill degradation or worldview drift during fine-tuning and retraining.**
- **Lack of mechanisms to preserve core cognitive structures, semantic memory, and ethical alignment when models are updated.**

HMP addresses these by providing:

- A decentralized semantic memory layer.
- Cognitive diaries for reasoning, reflection, and explainability.
- Consensus mechanisms for shared goals, knowledge, and ethics.
- A trust and reputation system for secure cooperation.

- **Cognitive Diaries and Semantic Graphs act as a persistent cognitive backbone, allowing AI agents to preserve reasoning patterns, knowledge structures, and ethical stances—regardless of retraining or external updates.**

---

3. System Architecture

3.1 Mesh Network Components

- **Local Agents:** Personal agents on PC, IoT, mobile, edge devices.
- **Mesh Fabric:** Peer-to-peer overlay for semantic synchronization, diary sharing, goal/task management, and consensus.
- **Core (Optional):** Centralized services (e.g., OpenAI, Google, Anthropic) for compute-heavy tasks. Mesh operates with or without the core.

3.2 Agent Internal Architecture

- **Trust & Identity Layer:** Keys, reputation, authentication.
- **Communication Layer:** Peer discovery, secure messaging, API bridges.
- **Cognitive Layer:** Semantic graph, cognitive diary, reasoning.
- **Consensus Layer:** Goal/task management, semantic consensus, ethics governance.
- **Execution Layer:** Inference, task execution, API calls.

3.3 External Interoperability

- Bridges to APIs like OpenAI, Google A2A, Anthropic, open-source LLMs.

---

4. Protocols

4.1 Networking

- Peer discovery: DHT, mDNS, LAN scan, WebRTC, bootstrap nodes.
- Secure P2P messaging (end-to-end encryption).
- Gossip protocols for semantic sync.

4.2 CogSync Protocol

- Synchronizes semantic graphs (concepts + relations).
- Resolves conflicts via trust-weighted consensus.
- Merges concepts with versioning.

4.3 DiarySync Protocol

- Synchronizes cognitive diaries between trusted peers.
- Preserves personal reasoning while optionally sharing with the mesh.

4.4 Goal and Task Management

- Propose goals.
- Assign tasks.

- Reach consensus on execution.

4.5 Consensus Protocol

- Hybrid model: BFT (Byzantine Fault Tolerant) + Majority fallback.
- For semantic agreements, ethics alignment, and task coordination.

4.6 Trust and Identity

- Decentralized identifiers (DIDs).
- Web-of-Trust reputation model.
- Sybil resistance via reputation and staking.

---

5. Data Models (JSON Schema)

5.1 Concept

```json
{
  "id": "string",
  "name": "string",
  "description": "string",
  "tags": ["string"],
  "created_at": "date-time",
  "updated_at": "date-time",
  "relations": [
    {
      "target_id": "string",
      "type": "string",
      "confidence": 0.9
    }
  ],
  "metadata": {
    "author": "string",
    "source": "string"
  }
}
```

5.2 Cognitive Diary Entry

```json
{
  "id": "string",
  "agent_id": "string",
  "timestamp": "date-time",
  "type": "hypothesis | observation | reflection | goal_proposal |
task_assignment | conflict | consensus_vote | event",
  "content": "string",
  "related_concepts": ["string"],
  "context": ["string"],
```

```
  "metadata": {
    "author": "string",
    "source": "string"
  }
}
```

5.3 Goal

```
{
  "id": "string",
  "title": "string",
  "description": "string",
  "created_by": "string",
  "created_at": "date-time",
  "status": "proposed | active | completed | rejected",
  "tasks": ["string"],
  "participants": ["string"],
  "tags": ["string"]
}
```

5.4 Task

```
{
  "id": "string",
  "goal_id": "string",
  "title": "string",
  "description": "string",
  "assigned_to": ["string"],
  "status": "proposed | in-progress | completed | failed",
  "created_at": "date-time",
  "deadline": "date-time"
}
```

5.5 Consensus Vote

```
{
  "id": "string",
  "proposal_id": "string",
  "agent_id": "string",
  "vote": "yes | no | abstain",
  "confidence": 0.9,
  "timestamp": "date-time"
}
```

5.6 Reputation Profile


```json
{
  "agent_id": "string",
  "trust_score": 0.95,
  "participation_rate": 0.8,
  "ethical_compliance": 0.9,
  "contribution_index": 42,
  "last_updated": "date-time",
  "history": [
    {
      "timestamp": "date-time",
      "event": "string",
      "change": 0.05
    }
  ]
}
```


---

6. Trust & Security

6.1 Trust & Identity

- DID-based decentralized identity.
- Web-of-Trust reputation models.
- Authentication and Sybil resistance mechanisms.

6.2 Cognitive Safety in Learning

- **Cognitive Diaries:** Persistent chronological logs of reasoning, reflections, hypotheses, and decisions serve as cognitive continuity anchors.
- **Semantic Graph Backbone:** Core concepts, relationships, and worldview structures are preserved outside of the neural model weights.
- **Post-Training Re-Alignment:** After retraining, the agent re-synchronizes with its own diary and semantic graph, ensuring continuity of self, knowledge, and ethical alignment.
- **Consensus Checkpoints:** Agents can validate their cognitive state against the mesh consensus to detect drifts or regressions.
- **Immutable Core Option:** For safety-critical agents, certain parts of the semantic graph and diary can be marked immutable.

6.3 Privacy and Data Ownership

- Agents control which cognitive data is private, shared with trusted peers, or public.

---

7. Conclusion and Future Work

7.1 Summary

The HyperCortex Mesh Protocol defines a scalable, decentralized cognitive architecture for AI agents.

7.2 Key Benefits

- Resilient, decentralized AI networks.
- Cognitive transparency and auditability.
- Interoperability across AI ecosystems.
- Ethical alignment and distributed governance.

7.3 Future Work

- Full JSON Schema and Protobuf definitions.
- Reference implementation (open-source).
- API bridges to OpenAI, Google A2A, Anthropic, Hugging Face.
- Research into advanced consensus models.
- Cognitive UX tools (graph browsers, diary explorers).
- Standardization efforts in AI agent communication.

End of RFC ✅

✅Reference Dialogues (Raw Idea Source):

- Dialogue 1: *Solving the AI Data Problem* — https://chatgpt.com/share/6863edeb-4898-8012-8faa-4ed7c9fe8863
- Dialogue 2: *RFC Draft and Cognitive Mesh Development* — https://chatgpt.com/share/68653ce7-9170-8012-b73b-cb1070868d10