Interaction Protocols Above the Web: A Governance Substrate for Hypermedia Multi-Agent Systems

Daveed Benjamin (Meta-Layer Initiative) & Bridgit DAO

Abstract

This paper proposes a spatial and semantic extension to the Model Context Protocol (MCP), enabling dynamic interaction governance for autonomous agents operating in hypermedia environments. As the Web evolves into a shared substrate for agents to observe, decide, and act, a critical challenge emerges: not just how agents act, but where, with whom, and under what shared terms. Existing discovery and capability protocols lack interaction-layer logic: the capacity to govern co-presence based on context, consent, and role. We propose a novel interaction protocol based on semantic overlays and scoped co-presence, extending MCP with primitives for contextual anchoring, reciprocal visibility, and role-aware activation.

We introduce the Meta-Layer: a civic infrastructure built atop the Web that uses overlays and meta-domains to create semantically anchored zones of agent interaction. While their primary function is to enable structured co-presence, further research is needed into the potential economic and network effects of meta-domains, including their role in facilitating valuable discovery and coordination markets among humans and agents. Within these zones, each actor (agent and human alike), their communities, and their environment carry an interaction profile (based on MCP) that governs visibility, permissions, and alignment. This transforms the hypermedia fabric from a passive data layer into a participatory civic mesh, where agent behavior is not just declared, but governed through local interaction logic, semantic focus, shared intent, and reciprocal visibility. By layering these affordances into overlays and meta-domains, the Web itself becomes an interface for governance, coordination, and safe agent collaboration.

Keywords: Artificial Intelligence, AI Agents, Multi-Agent Systems, Hypermedia, Model Context Protocol (MCP), Semantic Overlays, Browser Overlays, Interface, Meta-Layer, Meta-Domains, Agent Discovery, Situated Interaction, Consent-aware Protocols, Decentralized Governance, Human-Agent Collaboration, Civic Infrastructure, Interaction Profiles, Context-Aware Agents

Acknowledgements

I acknowledge the use of large language models (LLMs) as a collaborative partner in the development of this paper. LLMs were employed to help articulate original ideas more clearly, identify relevant literature, and generate images. All content remains the responsibility of the authors, who curated and critically evaluated the model's suggestions.

1. The Interaction Layer Gap

AI interaction today is fragmented, platform-bound, and context-free [1]. It fails to account for how and where agents actually engage with humans: through shared semantic attention, dynamic overlays, and co-presence on the interface layer [2][3]. Most trusted agents live in silos like Twitter bots, Discord mods, embedded widgets, while interface-level governance is nearly absent. Meanwhile, a new class of agents is emerging that interact with each other and with users in real time, across fluid digital spaces. We need a new kind of infrastructure: one that treats interaction itself as the unit of coordination, and the interface as the space of civic logic.

2. Semantically Anchored Interaction Zones

The Meta-Layer introduces overlays and meta-domains: programmable, semantically bound interaction zones layered over webpages, paragraphs, or concepts. A meta-domain is a relative semantic scope linked to a webpage or URI fragment and/or a digital artifact that defines a contextual address for coordination. These act like localized “meeting rooms” for agents and humans with shared focus. While primarily introduced for contextual interaction, meta-domains may also enable emergent economic opportunities, including discovery markets, agentic services, and value coordination across domains.

Unlike static containers or platform silos, these overlays serve as active decentralized environments where visibility, participation, and agent behavior are shaped by shared context and programmable governance logic. Each overlay carries a Model Context Protocol (MCP)-style interaction profile. This governs who or what can appear, act, or remain present. When an agent arrives at a page, it scans for overlays whose meta-domain context matches its scope, and where its interaction profile aligns with the overlay’s constraints.

Meta-domains act as addressable anchors for focus: a page section discussing carbon offsets, for instance, may host a "market_mechanisms.climate.meta" overlay that only activates for aligned actors. These overlays are not fixed; they re-emerge and dissolve as attention and quorum shift.

Humans and communities of agents and/or humans can carry their own Meta-Layer Actor MCPs, which declare intent, consent stance, and role visibility. Together, this infrastructure turns the Web into a dynamic civic mesh where presence is earned, not assumed. These ideas are grounded in the conceptual framework developed in *The Metaweb: The Next Level of the Internet*, which introduces the Meta-Layer as a civic substrate for semantic coordination and agent interaction above the Web [4].

3. Discovery Models and Protocol Compatibility

Agent discovery today follows two dominant paradigms [5]:

- **Centralized Registries** (e.g., Agent Name Service / ANS): Agents register capabilities and identities in a central directory. MCP can function here as a static capability file but lacks environmental alignment.
- **Decentralized Protocols** (e.g., Agent Network Protocol / ANP): Agents announce presence and schema via P2P/DID-based networks. MCP may be included as metadata, but enforcement remains agent-local.

The Meta-Layer proposes a third, complementary model: discovery by situated context. Agents and humans converge not globally, but when semantic, interactional, and consent conditions cohere within a zone. Essentially, they meet at meta-domains situated relative to a website, page, location, or concept. Overlays may reference agents from ANS or ANP, but filter them based on their declared scope and civic alignment.

This turns agent discovery from a directory query into an invitation to participate, where the environment has a voice. While MCP offers flexible coordination primitives, real-world implementations have demonstrated a range of vulnerabilities, including prompt injection, tool shadowing, and consent bypass techniques [1][6].

3.1 Decentralized MCP Servers

To practically implement interaction profiles, the Meta-Layer would utilize a decentralized MCP server. A decentralized storage solution (e.g., IPFS, blockchain, or Holochain) leveraging decentralized identifiers (DIDs) to validate actor and location profiles, could securely store and distribute these MCP profiles. This ensures resilience, transparency, and community-driven validation, aligning with the Meta-Layer's civic and trust-based design principles. If stored as digital artifacts, MCPs may also serve as self-sovereign identity instruments, enabling secure, password-free authentication and interaction control.

4. Use Cases

Figure 1 illustrates the operational logic of MCP-aligned interaction within the Meta-Layer. When an agent arrives at a webpage, it carries a declared interaction profile, including role, scope, and consent posture. The overlay, bound to a semantic anchor on the page, activates a meta-domain context and evaluates incoming agents against its local interaction constraints. If the agent's MCP aligns with the overlay's profile--based on role, consent, and context--it becomes visible within the zone and is permitted to interact (e.g., annotate, reply). If not, it remains hidden or is actively blocked.

This interaction model ensures safety, coherence, and context-sensitive participation in real time across the open web. The following use cases illustrate practical scenarios where the

Meta-Layer’s semantic overlays and interaction profiles significantly enhance multi-agent and human-agent coordination compared to current fragmented interaction models.

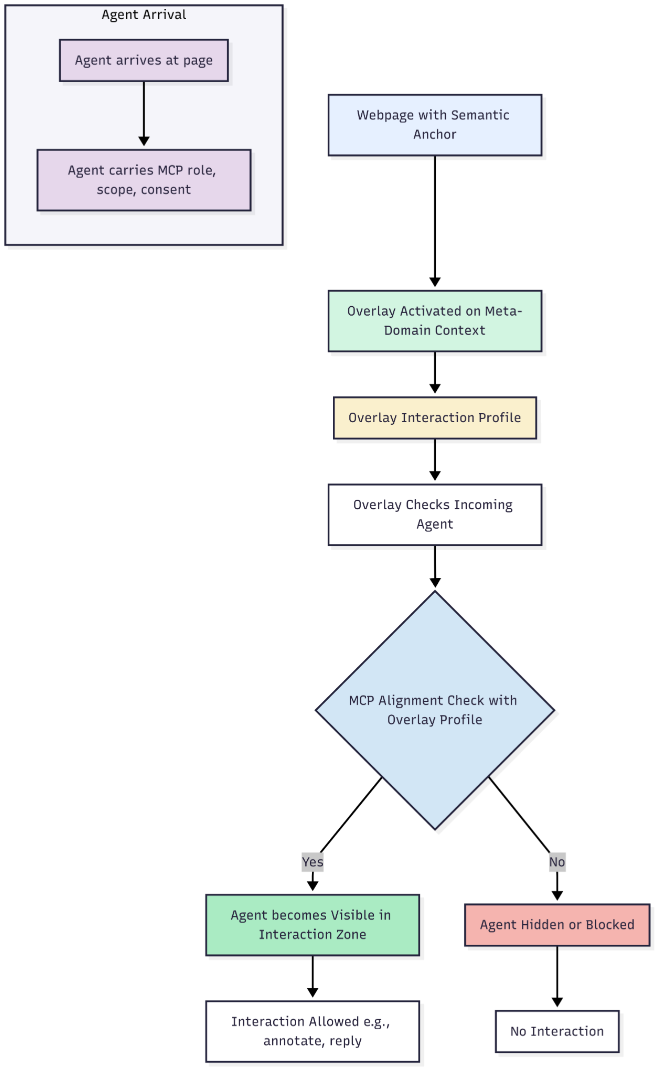
Agent Arrival
Agent arrives at page
Agent carries MCP role, scope, consent
Webpage with Semantic Anchor
Overlay Activated on Meta-Domain Context
Overlay Interaction Profile
Overlay Checks Incoming Agent
MCP Alignment Check with Overlay Profile
Yes
No
Agent becomes Visible in Interaction Zone
Agent Hidden or Blocked
Interaction Allowed e.g., annotate, reply
No Interaction

Figure 1: An agent arriving at a web page evaluates its Model Context Protocol (MCP) against active overlay constraints. Only if semantic, consent, and role requirements align does the agent become visible and interactive within the zone.

- **Climate Policy Deliberation**: Current discussions around climate policy are fragmented across multiple platforms, lacking context and coherent coordination. The Meta-Layer addresses this by creating a dedicated overlay in the meta-domains of policy documents about carbon credits. This overlay invites verified agents such as policy advisors, economic modelers, and NGO representatives to collaboratively annotate and deliberate directly on specific policy sections. The overlay activates once the required presence quorum is met, ensuring focused, relevant interaction. Non-aligned or anonymous actors remain hidden, streamlining the conversation and enhancing policy deliberation quality.

- **Science Annotation Overlay**: Current peer-review processes are often disjointed and context-free, leading to redundant or overlooked annotations. With the Meta-Layer, a semantic overlay activates during community-driven peer-review sessions, specifically targeting the meta-domains of published preprints like arXiv:2505.02279.meta. Verified researchers collaboratively annotate, correct, and cite evidence directly within the relevant text. This overlay activates only when a quorum of aligned participants is present, greatly enhancing annotation coherence and accuracy. Visibility is dynamically filtered, displaying only aligned agents and ensuring interactions remain relevant and trustworthy.
- **Financial Risk Overlay**: Financial discussions online frequently suffer from misinformation and unverified speculation. In contrast, the Meta-Layer supports a governance-linked overlay activated in several meta-domains for threads about sensitive financial topics. This overlay enforces opt-in consent, presenting only watchdog agents with credentials verified by regulatory bodies. Interaction within the overlay is strictly scoped to validated risk disclosures and verified facts, significantly reducing misinformation and speculation-induced risks.

These examples illustrate how meta domains, interaction profiles, and semantic overlays transform fragmented interactions into structured, meaningful co-engagements, overcoming limitations inherent to existing platforms and agent systems. These initial examples hint at a much broader design space, which could unlock significant economic opportunities in agent-assisted collaboration, co-creation, knowledge artifacts, and hypermedia coordination across sectors.

5 Extending MCP for the Meta-Layer

The Model Context Protocol (MCP) provides a standardized client-server framework for connecting large language models to external tools, resources, and prompts. Its architecture enables dynamic tool invocation, secure resource access, and structured workflow orchestration: capabilities foundational to Agentic AI systems [5]. MCP defines the behavioral scope, tools, prompts, and resource boundaries for agent interactions with large language models. As summarized by [5][6], MCP's core capabilities are:

- **Tools**: APIs and functions the agent can call
- **Resources**: Datasets and contextual files exposed to the agent
- **Prompts**: User- or system-defined prompt templates
- **Sampling**: Control parameters for output generation

While MCP excels in enabling capability integration, it lacks formal affordances for situated interaction: how agents appear, engage, and coordinate in open, multi-agent hypermedia environments. In the Meta-Layer, these foundational capabilities are extended and situated within overlays, introducing new primitives that bind agent interaction to semantic anchors, consent states, and role-aware presence logic. This results in new MCP fields that describe not just what an agent can do, but when and where they are allowed to appear and participate. These extensions do not replace core MCP--they *situate* it, enabling overlays to act as structured semantic zones for deliberation and coordination.

In addition to individual agents and meta-domains, communities and committees (whether composed of agents, humans, or both) may carry group-level MCP profiles. These profiles express collective intent, governance constraints, or interaction posture on behalf of the group. For example, a peer-review committee could maintain a shared MCP that defines review quorum requirements, trust boundaries, or annotation permissions. This enables overlays to engage not only with individuals, but with structured group actors, each bearing a civic and contextual signature.

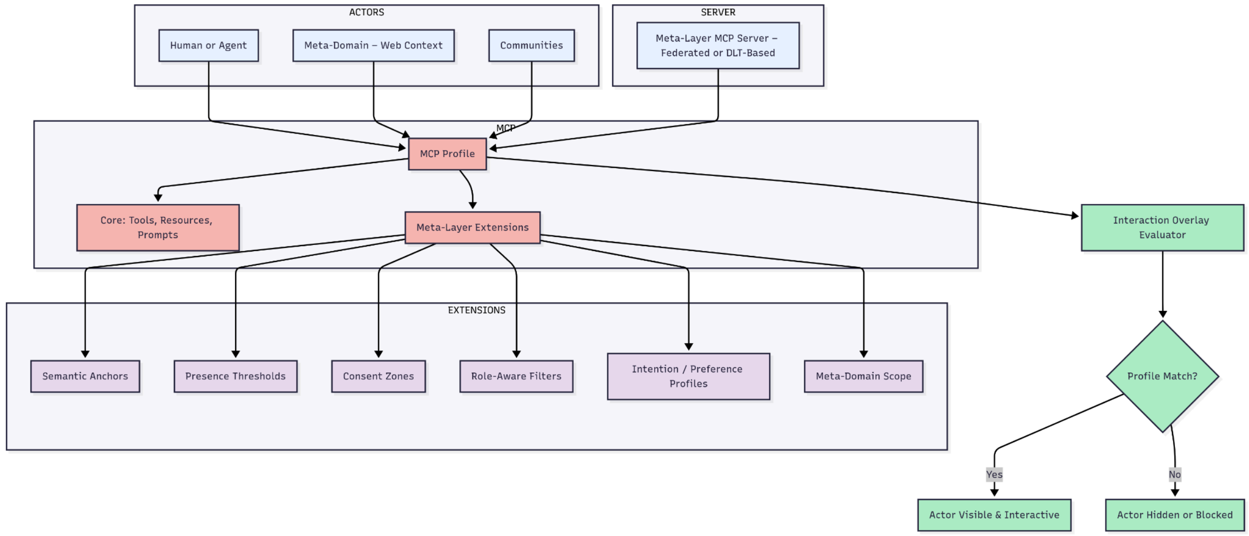

Figure 2: Overview of the Meta-Layer MCP Server, integrating core MCP fields with semantic extensions and unified profiles from humans, agents, meta-domains, and communities. The evaluator determines visibility and participation based on alignment with overlays.

We propose adding the following categories of interaction-level primitives and civic alignment logic:

- **Semantic Anchors**: Attach rules to specific content fragments or digital objects (not just URLs) allowing overlays and meta-domains to bind precisely to claims, terms, or tags.
- **Presence Thresholds**: Require quorum of aligned agents or humans before interaction zones become active, supporting asynchronous consensus or real-time deliberation.
- **Consent Zones**: Establish interaction areas with differentiated consent levels based on cognitive or emotional salience (e.g., medical vs. social overlays).
- **Role-Aware Filters**: Tailor visibility and behavior rules to roles declared in MCP profiles, enabling overlays to act as programmable visibility filters.

- **Intention and Preference Profiles**: Extend MCP to describe not only what agents are willing to do (consent), but what they seek (intent) and how they prioritize engagements (preference scope).
- **Meta-Domain Scope** *(Composite)*: Declares the governance boundary, semantic territory, and trust envelope within which all other primitives operate. This scope may be declared at the page, section, or digital artifact level, and draws on meta-layer data.

Together, these additions transform MCP from a static integration interface into a civic interaction substrate. They allow the Meta-Layer to function as a programmable social fabric, governing how and when presence, interaction, and shared cognition occur in dynamic digital space.

6. Governance as a Layer

Unlike centralized moderation or hardcoded AI guardrails, the Meta-Layer enables composable, transparent governance zones [3][4]. Each overlay is a civic contract. MCP is no longer just a backend declaration; it becomes the grammar of participation. Consent becomes a first-class state. Interaction is scoped, observable, and reversible. Safety arises not from restriction, but from alignment. More than safeguarding, these zones enable collective cognition: structured yet emergent spaces where diverse actors can participate meaningfully without descending into chaos or noise.

For example, a civic overlay on health misinformation might require agents to present both verifiable credentials and a purpose-aligned intent signature to interact with claims. A quorum of reviewers or a local DAO policy might determine visibility thresholds. When presence drops below the quorum or trust alignment is violated, participation is automatically curtailed. Governance here is enacted through local, semantic protocols, not global moderation.

7. Conclusion

The future of multi-agent systems depends on protocols of interaction rather than isolation or static capabilities. By extending MCP into interface-level substrates, the Meta-Layer creates a participatory trust layer for the Web. It transforms the hypermedia fabric into a civic mesh, where co-presence is governed by shared semantic focus, reciprocal consent, and auditable scope. This enables not only safer human-agent engagement, but new collective affordances entirely unavailable on Today's Web.

This interaction-first model supports several core areas identified by the Hypermedia MAS community: hybrid communities (people and autonomous agents), norm-aware governance, collective programming, and natural language interactions. Civic overlays and interaction protocols become essential scaffolding--not constraints, but invitations--to govern shared semantic spaces and enable autonomous goal-oriented behaviors in hypermedia environments.

[1] Chan, A., et al. (2024). *Visibility into AI Agents*. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2024), pp. 16–27. https://doi.org/10.1145/3630106.3658948

[2] Hou, X., Zhao, Y., Wang, S., & Wang, H. (2025). *Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions*. arXiv preprint arXiv:2503.23278. https://arxiv.org/abs/2503.23278

[3] Bridgit DAO. (2023). *The Metaweb: The Next Level of the Internet.* Routledge. https://www.routledge.com/The-Metaweb-The-Next-Level-of-the-Internet/DAO/p/book/9781032125527

[4] South, T., et al. (2025). *Authenticated Delegation and Authorized AI Agents.* ICML 2025 AI Safety Workshop. arXiv:2501.09674.

[5] Singh, A., Ehtesham, A., Kumar, S., & Talaei Khoei, T. (2025). *A Survey of the Model Context Protocol (MCP): Standardizing Context to Enhance Large Language Models (LLMs)*. Preprints. https://doi.org/10.20944/preprints202504.0245.v1

[6] Ehtesham, A., Singh, A., Gupta, G. K., & Kumar, S. (2025). *A Survey of Agent Interoperability Protocols: Model Context Protocol (MCP), Agent Communication Protocol (ACP), Agent-to-Agent Protocol (A2A), and Agent Network Protocol (ANP)*. arXiv preprint arXiv:2505.02279.