

Communication and Coordination in the Generic User Framework

Paul Hanchett with assistance from Anthropic Claude

Introduction

The Generic User approach to AI-browser interaction proposes a dual-AI architecture consisting of a Browser Interaction Specialist and a Domain Specialist. This paper explores the communication mechanisms between these agents and considers the potential for broader multi-agent coordination within this framework.

Agent Communication Mechanisms

The communication between the Specialized (Domain) agent and the Generic (Browser Interaction) agent presents significant design considerations. We identify three primary approaches for this inter-agent communication:

1. Natural Language Interface

Using natural language (such as English) as the communication medium between agents provides flexibility and expressiveness. This approach leverages the inherent capabilities of Large Language Models and allows for nuanced instruction.

Advantages:

- Intuitive for humans to understand and audit
- Highly flexible for expressing complex intentions
- Leverages existing LLM capabilities

Challenges:

- Potential for ambiguity and misinterpretation
- Historical examples from military communications and legal documents demonstrate how natural language can be misunderstood over time
- May require additional context to resolve ambiguities

2. Structured Task Representation

A formalized, structured approach using defined schemas for representing tasks and actions provides greater precision but less flexibility.

Example:

json

```
{
  "task": "login",
  "parameters": {
    "username": "user123",
    "password": "[MASKED]",
    "remember_me": true
  },
  "success_criteria": {
    "expected_destination": "dashboard",
    "expected_elements": ["welcome_message", "user_menu"]
  }
}
```

Advantages:

- Precise and unambiguous
- Aligns with training data structure
- Failed interactions can directly generate new training examples
- Easier to validate and parse programmatically

Challenges:

- Less flexible for handling novel or unforeseen situations
- More difficult for humans to interpret quickly
- Requires maintaining and extending a schema

3. Hybrid Approach

A combination of structured commands with natural language explanations provides both precision and flexibility.

Advantages:

- Structured components for precision
- Natural language for handling edge cases and reasoning
- Better human oversight capabilities

Challenges:

- More complex implementation
- Potential inconsistencies between structured and unstructured components

Direct vs. Mediated Communication

An important architectural consideration is whether applications should communicate directly with the Generic agent or only through the Specialized agent.

Mediated Communication:

- Cleaner separation of concerns
- Better encapsulation of domain knowledge
- Additional latency in processing
- Specialized agent can filter and translate application requirements

Direct Communication:

- More efficient for certain tasks
- Risks blurring domain boundaries
- May bypass important domain-specific constraints
- Requires Generic agent to understand application contexts

Multi-Agent Coordination

Extending beyond the dual-AI architecture raises questions about coordination among multiple agents. Potential challenges include:

Resource Contention

When multiple agents operate simultaneously:

- Bandwidth limitations in communication channels
- Contention for shared resources (browser control)
- Computational resource allocation
- Memory management across agents

Conflict Resolution

Agents competing for actions or making contradictory decisions require resolution mechanisms:

- Priority-based systems

- Task allocation protocols
- Hierarchical decision structures
- Consensus algorithms

Domain Boundary Disputes

Clear domain separation is preferable to conflict resolution:

- Well-defined responsibilities for each agent
- Established protocols for inter-agent handoffs
- Clear success criteria for tasks
- Domain-specific expertise without overlap

The Memory Agent Concept

A potentially valuable extension to the Generic User framework is a dedicated Memory Agent that:

- Monitors all inter-agent communications
- Maintains contextual awareness across sessions
- Stores significant information with temporal relevance
- Provides relevant historical context without disrupting primary interactions

This Memory Agent would serve as a form of institutional memory, recalling previous successful interactions with interfaces and providing context to both Specialized and Generic agents when relevant.

Conclusion

The communication mechanisms between agents in the Generic User framework significantly impact its flexibility, reliability, and extensibility. While natural language offers intuitive flexibility, structured representations align better with training methodologies. A hybrid approach may provide the best balance for most implementations.

As the framework expands to incorporate multiple agents, coordination mechanisms become increasingly important. Preventing conflicts through clear domain separation is preferable to complex resolution systems. The addition of a Memory Agent could enhance the system's capabilities by providing contextual awareness across interactions.

Future work should explore practical implementations of these communication approaches and evaluate their effectiveness in real-world browser automation scenarios.