How to Map Human Activities onto AI Models

Abstract

This document examines the architectural problem of mapping human activities onto AI models in a way that preserves semantic coherence, limits unintended authority, and avoids the silent propagation of conceptual inconsistency. It argues that direct role replication is a category error, that effective mapping requires prior human self-modeling, and that AI agents must be defined in terms sympathetic to their actual operational properties rather than anthropomorphic analogies. The objective is to establish conditions under which AI agents can participate in complex workflows without amplifying ambiguity or eroding systemic integrity.

1. Problem Statement

As AI systems are introduced into knowledge-intensive workflows, they are frequently framed as replacements or augmentations for human roles. Common examples include “AI business analysts,” “AI architects,” or “AI reviewers.” These framings presume that human activities can be transferred wholesale to non-human actors.

In practice, this assumption fails. Human roles are socially constituted, implicitly bounded, and continuously repaired through tacit judgement. AI agents, by contrast, operate exclusively over explicit representations and cannot participate in informal correction.

The result is not immediate malfunction but a gradual loss of semantic control: outputs appear locally coherent while becoming globally misaligned across domains. This failure mode is structural rather than accidental.

The core architectural question is therefore not how to make AI agents more human-like, but how to **map human activities onto AI models without importing the ambiguities embedded in human roles**.

2. Human Roles as Composite Structures

Human organisational roles are not atomic. They conflate several distinct functions:

- information transformation
- decision-making authority
- accountability for outcomes
- social coordination and negotiation

These functions coexist within a single role because humans can tolerate ambiguity, repair misunderstandings, and adapt behaviour contextually.

AI agents cannot perform this repair. Any attempt to model an AI agent directly on a human role therefore embeds hidden assumptions about authority, intent, and judgement that the agent does not possess.

This explains why role-based AI agents often appear competent while exceeding their legitimate scope. The role abstraction conceals which aspects of human activity are informational and which are irreducibly social or decisional.

3. Essential Properties of AI Agents

An AI agent is best understood as a constrained transformation system operating over symbolic artefacts.

It exhibits the following properties:

- It operates exclusively on representations rather than on reality or intent.
- It has no intrinsic goals, authority, or accountability.
- It cannot detect conceptual incoherence unless incoherence is explicitly encoded.
- It produces outputs that are locally plausible within a supplied context.
- It does not possess epistemic awareness of correctness or error.
- It scales pattern application rather than judgement.

These properties are not limitations to be overcome. They are defining characteristics that must shape architectural design.

Any mapping that ignores these properties will produce agents that appear useful while silently degrading semantic integrity.

4. The Necessity of Human Self-Modeling

Before human activities can be mapped onto AI models, humans must first model their own activities coherently.

In most organisations, this modeling has never been made explicit. Roles are understood through convention, apprenticeship, and social feedback rather

than through formal definition. Outputs are evaluated pragmatically rather than against declared semantic contracts.

Humans compensate for this informality by joining dots across documents, meetings, and systems. This compensation is local, temporary, and unrecorded.

AI agents remove this safety net. They accept misaligned inputs without protest and propagate their consequences across linked domains. As scale increases, humans lose the ability to perceive mismatches between domains of production.

The introduction of AI therefore exposes a pre-existing architectural weakness: **human activities were never sufficiently specified to be safely externalised**.

5. Mapping as Decomposition, Not Replication

Effective mapping does not proceed from roles to agents. It proceeds from **activities to functions**.

The correct unit of mapping is not “what a person is,” but “what transformations they perform.” This requires decomposing human activities into:

- informational transformations
- decision points
- authority boundaries
- responsibility assignments

Only the first of these can be directly mapped to AI agents.

The mapping is therefore inherently many-to-many:

- a single human role decomposes into multiple agent-compatible functions,
- and a single agent function may support multiple human roles.

This asymmetry is not a defect. It reflects the fact that human roles were never designed as computational abstractions.

6. Training Inputs as Semantic Contracts

Once activities are decomposed, the question of training input becomes precise.

Artefacts suitable for training or operational context must:

- make distinctions explicit rather than implicit,
- encode intent and constraint rather than outcome alone,
- be consumable without inference or background knowledge,
- declare scope, limits, and uncertainty.

Unstructured text, informal explanation, and post-hoc narrative may be valuable for human understanding, but they do not establish binding meaning for an AI agent.

Training on such artefacts teaches surface regularities while leaving semantic authority undefined.

This produces agents that are fluent but ungrounded, and systems that are productive but incoherent.

7. Authority, Ownership, and Responsibility

AI agents do not own decisions. Any apparent decision is enacted only when a human system accepts the output.

Architectures must therefore make authority explicit:

- what the agent may generate,
- what it may not decide,
- and where human intervention is mandatory.

Failure to define these boundaries results in authority leakage, where outputs are treated as decisions without accountability.

This is not an AI ethics problem. It is a systems design problem.

8. Conclusion

Mapping human activities onto AI models is not an exercise in imitation. It is an exercise in clarification.

AI agents do not reduce the need for conceptual clarity; they expose its absence and scale its consequences. Organisations that introduce AI without first modeling their own activities will experience increasing internal inconsistency masked by fluent output.

A sympathetic mapping approach begins by acknowledging what AI agents are in fact, decomposing human activities accordingly, and treating artefacts as semantic contracts rather than narrative conveniences.

Only under these conditions can AI agents participate safely and effectively in complex workflows without becoming engines of untraceable ambiguity.