

Intent Resolution as an External System in AI-Assisted Workflows

Alexander Sucala

Independent Researcher
(2026)

Abstract

Human interaction with AI systems is commonly mediated through natural language prompts. In many deployed architectures, these prompts are treated as both expressions of intent and executable instructions. This conflation introduces ambiguity, scope drift, security risk, and irreproducible behavior, particularly as systems accumulate memory, state, and long-term objectives.

This paper argues that intent resolution must be externalized from large language models (LLMs) and governed as a deterministic system component. Rather than permitting models to infer intent implicitly from conversational context, AI systems should translate human input into structured, bounded task representations using explicit state, memory, and policy inputs. The LLM then operates only within the constraints of this resolved intent.

This work defines architectural requirements for external intent resolution without prescribing specific algorithms or implementations, establishing necessary conditions for reliable, auditable, and scalable AI-assisted workflows.

1. Introduction

Natural language is an effective interface for expressing human goals, but it is an unreliable medium for defining executable system behavior. In many AI-assisted workflows, user prompts are passed directly to LLMs, which are then expected to infer intent, determine scope, recall relevant context, and generate appropriate responses.

This design assigns interpretive authority to a probabilistic component and collapses multiple system responsibilities into a single conversational interaction. As systems scale in complexity and autonomy, this approach becomes increasingly fragile.

This paper contends that **intent resolution is a system responsibility, not a model capability**, and must be handled externally to ensure correctness, security, and reproducibility.

2. The Prompt–Intent Conflation Problem

A core failure in many AI systems is the assumption that a user’s prompt is a complete and executable representation of intent.

In practice, human input is often:

- underspecified,
- context-dependent,
- emotionally framed, or
- ambiguous with respect to scope and authority.

When LLMs are tasked with resolving these ambiguities internally, systems inherit the model’s probabilistic interpretation as authoritative intent.

This leads to:

- inconsistent behavior across sessions,
 - unintended scope expansion,
 - vulnerability to prompt injection or manipulation, and
 - difficulty auditing why a system behaved as it did.
-

3. Intent as a First-Class System Object

This paper proposes treating intent as a **first-class system object**, distinct from raw user input and distinct from model output.

Under this framing:

- user input is an **expression**, not an instruction,
- intent is a **resolved representation** derived from multiple contextual inputs, and
- execution occurs only within the bounds of the resolved intent.

Intent resolution is therefore a transformation process governed by system rules rather than linguistic inference.

4. Externalized Intent Resolution

Externalized intent resolution separates interpretation from generation.

In such systems:

- user input is processed by deterministic logic outside the LLM,
- relevant context may include explicit state, historical interactions, goals, or policy constraints, and
- the output of this process is a structured task description provided to the LLM.

The LLM is not permitted to redefine intent, expand scope, or infer authority beyond what is explicitly provided.

This separation ensures that intent remains stable, inspectable, and auditable.

5. Memory and State as Inputs, Not Thoughts

Many AI systems treat memory as an emergent property of conversation. Prior outputs, conversational turns, or internal embeddings are implicitly relied upon to guide future behavior.

This approach obscures causality and complicates validation.

In contrast, reliable systems treat memory and state as **explicit inputs** to intent resolution rather than internal model cognition. Memory does not influence behavior unless it is deliberately selected and supplied by the system.

This design preserves control over relevance, scope, and authority without requiring disclosure of selection mechanisms.

6. Deterministic Translation and Bounded Execution

Once intent is resolved, it may be translated into a form suitable for model execution. This translation must be:

- deterministic,
- bounded by explicit constraints, and
- independent of model-internal reasoning.

The purpose of translation is not to optimize language quality, but to ensure that the model operates within an authorized and well-defined task space.

LLMs then function as probabilistic execution engines acting on a constrained input rather than as interpreters of human intent.

7. Security and Scope Control

External intent resolution provides significant security benefits.

By separating expression from execution:

- unauthorized scope expansion is prevented,
- sensitive actions require explicit authorization, and
- conversational manipulation cannot override system policy.

Intent resolution thus becomes a critical enforcement layer for AI systems operating in environments with real-world impact.

8. Scope and Limitations

This paper does not:

- define intent schemas,
- prescribe memory retrieval methods, or
- specify translation algorithms.

Its goal is to establish architectural requirements for intent governance rather than to describe implementation strategies.

9. Conclusion

As AI systems become more autonomous and context-aware, the distinction between human expression and executable intent becomes increasingly important. Treating prompts as executable instructions places interpretive authority in probabilistic components ill-suited for governance.

By externalizing intent resolution, treating intent as a first-class system object, and bounding model execution through deterministic translation, AI systems can achieve greater reliability, security, and auditability.

Ultimately, scalable AI systems will be defined not by how fluently they respond, but by how clearly and safely they understand what they are permitted to do.