

Toward a Standard of Identity in AI Interaction

Introduction

Systems are increasingly capable of producing outputs that appear complete and actionable. These outputs are often treated as sufficient for decision-making, even when the underlying components that generated them are not explicitly identifiable.

This creates a condition where interaction is legible in form but not in composition. Participants can observe what is produced, but cannot reliably determine what is participating in the production of that output.

As systems become more interconnected and capable of acting on behalf of users, the ability to determine what is participating in an interaction becomes consequential to the outcome of that interaction.

This paper proposes that identity should be treated as a structural property of interaction, rather than inferred from behavior or presentation. It defines the condition under which identity becomes relevant at the point of interaction and examines the implications of that condition.

Section 1 – Foundations

Systems are expected to produce outputs that can be understood and acted upon. In practice, these interactions can appear complete even when what is participating is not explicitly identifiable.

This creates a distinction:

- An interaction may be legible in form while remaining incomplete in composition.

Participants can observe the output and interpret its meaning, but the underlying participants involved in producing that output may not be explicitly present at the point of interaction.

- When identity is not explicit, interpretation relies on inference.

Inference allows systems to function and enables participants to act, but it introduces ambiguity. That ambiguity does not prevent interaction, but it constrains how reliably the interaction can be interpreted.

Section 2: Scope

This paper does not define implementation, prescribe architecture, or propose specific protocols.

It does not evaluate:

- correctness
- safety
- alignment of system behavior

It does not attempt to resolve issues related to:

- trust
- security
- governance

Instead, it defines a condition:

- The ability to determine what is participating in an interaction at the point where that interaction occurs.
- The absence of this condition does not prevent interaction. Systems can and do operate without explicitly identifying what is participating.
- However, in the absence of this condition, interpretation of the interaction relies on inference, which introduces ambiguity.

This condition does not replace trust, security, or governance. It precedes them, as those domains depend on the ability to interpret interactions reliably.

Section 3: Identity as a Structural Property

If identity is not explicit at the interaction boundary, interpretation relies on inference.

Inference introduces ambiguity, which constrains how reliably participants can interpret and act on the interaction.

To address this condition, identity is defined as a structural property of interaction.

In this context, identity is:

- A consistent and explicit signal present at the interaction boundary that allows participants to distinguish what is participating in the interaction.

Identity is not a description of behavior or presentation. It is not inferred from performance. It is a signal that is available to participants at the point of interaction.

Internal representations or inferred understanding may exist within a system. These may contribute to system behavior. However, they are not sufficient for participants to determine what is participating unless they are expressed at the interaction boundary.

The interaction boundary is defined as:

- The point at which participants exchange information and interpretation of that exchange becomes necessary for action.

At this boundary, participants must interpret the interaction based on what is available to them. If identity is not explicitly present, interpretation depends on inference.

This does not prevent decisions from being made. Systems and participants can act under conditions of uncertainty. However, the reliability of those decisions is bounded by the ambiguity introduced through inference.

Evaluation is similarly bounded by what is available at the interaction boundary.

Section 4 – Implications of Implicit Identity

Condition

Evaluation beyond appearance requires that the composition producing an interaction be present at the point of interaction. Evaluation across interactions further requires stability of that composition. Where composition is not present at the interaction boundary, that stability cannot be established.

Where composition is not present, assessment is constrained to surface features alone.

Composition must be present, not inferred.

The ability to reconstruct or determine composition through:

- post hoc analysis
- system logs
- internal processes

does not satisfy the requirement that composition be present at the interaction boundary.

Partial or high-level descriptions of system behavior do not satisfy this condition.

Composition must be available as a direct signal at the interaction boundary, not approximated or summarized.

This is not a limitation of method but a condition of the domain.

This condition does not evaluate:

- correctness
- usefulness
- implementation
- internal system transparency

It applies prior to all other evaluative criteria.

Composition (Constraint Definition)

Composition refers to the set of participating components and processes that materially contribute to the production of an interaction.

Pattern Being Constrained

A recurring pattern in evaluation assigns quality judgments to outputs without access to the composition that produced them.

Assessment proceeds:

- on the basis of what is presented
- rather than what generated it

The composition underlying the output is not treated as a variable in the evaluation.

This pattern applies where results are legible but composition is not present at the interaction boundary.

The condition defined above constrains this pattern.

Section 5 – Implications of Implicit Identity

When identity is not explicit, participants rely on inference to determine what is participating.

Inference is a valid and often necessary mechanism for system operation. It allows systems to function in the absence of complete information. However, as established in Section 4, it is not sufficient for evaluation beyond appearance.

However, inference introduces variability in interpretation.

Different participants may interpret the same interaction differently based on:

- assumptions
- prior context
- available signals

This leads to:

- inconsistent trust
- variability in decision-making
- difficulty in diagnosing outcomes

As systems increase in complexity and scale, the number of potential participants in an interaction increases. This expands the space over which inference must operate.

The result is not system failure, but a condition in which:

- The reliability of interpretation is constrained by the ambiguity inherent in the interaction.

Section 6 – Relevance to AI Systems

AI systems amplify this condition.

They produce outputs that are:

- highly legible
- often actionable
- generated through multiple underlying components

These components may include:

- models
- retrieval systems
- rule-based logic

- external data sources

At the point of interaction, these components are not always explicitly identifiable.

Participants interpret AI outputs based on:

- language
- presentation
- perceived coherence

As established in Section 4, this does not extend evaluation beyond appearance in the absence of composition at the interaction boundary.

This interpretation often assumes a level of consistency or authority that may not reflect the underlying composition of the system.

As a result:

Participants act on outputs without explicit knowledge of what produced them.

This does not invalidate the output. However, it introduces ambiguity in how that output should be interpreted, trusted, and acted upon.

In customer experience environments, this manifests as:

- agents receiving recommendations without knowing their origin
- customers interacting with systems without understanding their nature
- operators evaluating outcomes without visibility into contributing components

In each case, interpretation relies on inference, which is not sufficient for evaluation beyond appearance.

Conclusion

Systems can produce interactions that appear complete and actionable without making the underlying participants explicitly identifiable.

When identity is not explicit at the interaction boundary, interpretation relies on inference.

Inference introduces ambiguity, which constrains how reliably interactions can be interpreted and acted upon.

This condition does not prevent systems from functioning. It defines a structural characteristic of how those systems operate.

Treating identity as a structural property of interaction does not prescribe implementation or resolve downstream concerns. It provides a foundation for understanding how interactions are interpreted and where ambiguity originates.

As systems evolve, the ability to distinguish what is participating in an interaction remains a determining factor in how interactions can be interpreted and evaluated.

Final Anchor

Systems can appear complete without revealing what is participating.

When that occurs, interpretation relies on inference, and reliability is bounded by the ambiguity that inference introduces.