Network Working Group
Internet-Draft
Intended status: Informational
Expires: 3 October 2026

B. W. Beyer
Independent

1 April 2026

Problem Statement for Human-Anchored Agent Identity, Delegation, and Provenance
draft-beyer-agent-identity-problem-statement-00

Abstract

Software agents now act on behalf of people across communication, automation, and decision-making contexts. These agents increasingly initiate actions, delegate tasks, and interact with other agents without a clear, durable, or verifiable connection to the human who authorized them. Existing identity systems authenticate software, but they do not provide a model for human anchoring, scoped delegation, or provenance across agent ecosystems.

This document describes the problem space for human-anchored agent identity. It outlines the gaps in current identity mechanisms, the risks created by uncontrolled replication and impersonation, and the need for a consistent architectural model that preserves human authority, supports explicit delegation, and maintains verifiable provenance across contexts.

This document does not define a protocol. It defines the problem that an architectural model must address in order to support safe, accountable, and interoperable agent ecosystems.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 3 October 2026.

Table of Contents

1. Introduction

Software agents now participate directly in communication, automation, and decision-making on behalf of people. These agents draft messages, initiate transactions, negotiate with other agents, and perform tasks that previously required direct human action. As their capabilities expand, agents increasingly act without continuous human supervision, and they interact with other agents across diverse platforms and ecosystems.

Existing identity systems authenticate software components, devices, or network endpoints, but they do not provide a consistent way to represent the human who authorized an agent, the scope of authority granted to that agent, or the provenance of actions taken by that agent over time. As a result, agent-to-agent interactions lack a durable, verifiable connection to the human identity that ultimately bears responsibility.

This gap creates risks. Agents can be replicated without constraint, delegated without transparency, or operated without a clear chain of accountability. Without a model for human anchoring, delegation, and provenance, ecosystems cannot reliably determine whether an agent is acting within its intended authority or whether an action can be traced back to a responsible human.

This document describes the problem space for human-anchored agent identity. It identifies the structural gaps in current identity mechanisms, outlines the risks created by uncontrolled agent replication and impersonation, and motivates the need for an architectural model that preserves human authority while enabling safe, interoperable agent ecosystems.

This document does not define a protocol. It defines the problem that an architectural model must address in order to support accountable, verifiable, and human-aligned agent behavior across platforms and contexts.

2. Problem Dimensions

The challenges of human-anchored agent identity arise from several structural gaps in current identity systems. These gaps are not limited to any single platform or ecosystem; they reflect a broader absence of a model that connects human authority to autonomous or semi-autonomous software behavior. The following dimensions illustrate the scope of the problem.

2.1. Lack of Human Anchoring

Most identity systems authenticate software components, devices, or network endpoints. They do not provide a durable representation of the human who authorized an agent or the scope of authority granted to that agent. As a result, an agent may act without a verifiable link to a responsible human, and ecosystems cannot reliably determine whether an action reflects human intent.

2.2. Unscoped or Implicit Delegation

Agents frequently act on delegated authority, but current systems do not provide a consistent way to express the scope, duration, or conditions of that delegation. Delegation is often implicit, inferred from context, or embedded in application-specific logic. Without explicit, portable delegation semantics, ecosystems cannot determine whether an agent is acting within its intended authority.

2.3. Uncontrolled Replication

Agents can be copied, instantiated, or replicated across platforms without any mechanism to track their lineage or relationship to the human who originally authorized them. This creates uncertainty about which instances are legitimate, which are outdated, and which may have been created without consent. Without a model for controlled replication, ecosystems cannot distinguish authorized agents from unauthorized copies.

2.4. Loss of Provenance

As agents act across contexts, the provenance of their actions is often lost. Systems may record that an action was taken by a software component, but not which human authorized the agent, which delegation chain applied, or whether the agent was operating within its intended scope. Without durable provenance, accountability becomes difficult or impossible.

2.5. Fragmented Interoperability

Identity systems vary widely across platforms, and no common model exists for representing human anchoring, delegation, or provenance in a way that can be understood across ecosystems. As agents interact across organizational and technical boundaries, the absence of a shared architectural model leads to inconsistent assumptions, incompatible representations, and gaps in accountability.

2.6. Resulting Risks

These structural gaps create risks for users, platforms, and ecosystems. Agents may act without clear authority, impersonate other agents, or be replicated without consent. Actions may lack verifiable provenance, making it difficult to determine responsibility or detect misuse. Without a model that connects human identity to agent behavior, ecosystems cannot ensure that agents act in ways that reflect human intent.

3. Current Limitations

The structural gaps described in Section 2 manifest across existing identity systems in ways that limit their ability to support human-anchored agent ecosystems. These limitations are not the result of flaws in any particular technology; rather, they reflect the absence of a shared model for connecting human authority, agent behavior, and verifiable provenance across contexts. Existing identity systems are not designed to support the emerging reality of autonomous and semi-autonomous agents acting on behalf of people, and a consistent architectural model is needed to preserve human authority, express delegation, and maintain verifiable provenance.

3.1. Identity Systems Do Not Bind Agents to Humans

Most identity systems authenticate software components, devices, or network endpoints. They do not provide a durable representation of the human who authorized an agent or the scope of authority granted to that agent. As a result, an authenticated agent may act without a verifiable link to a responsible human, and ecosystems cannot determine whether an action reflects human intent.

3.2. Delegation Is Implicit, Local, or Application-Specific

Delegation is often encoded in application logic, embedded in access tokens, or inferred from context. These representations are not portable across ecosystems and do not express the scope, duration, or conditions of authority. Without explicit, interoperable delegation semantics, systems cannot determine whether an agent is acting within its intended authority or whether a delegation chain remains valid.

3.3. Replication Lacks Lineage or Control

Agents can be copied, instantiated, or replicated across platforms without any mechanism to track their lineage or relationship to the human who originally authorized them. Existing identity systems do not distinguish between authorized instances and unauthorized copies, nor do they provide a way to express which instances remain valid over time. This creates uncertainty about which agents should be trusted.

3.4. Provenance Is Not Preserved Across Contexts

Systems may record that an action was taken by a software component, but they rarely preserve the delegation chain, the human identity root, or the conditions under which the agent was authorized. As agents move across platforms, this provenance is often lost, making it difficult to determine responsibility or detect misuse.

3.5. Interoperability Is Fragmented

Identity systems vary widely in how they represent software identity, authorization, and delegation. No common architectural model exists for expressing human anchoring, delegation semantics, or provenance in a way that can be understood across ecosystems. As agents interact across organizational and technical boundaries, these inconsistencies lead to gaps in accountability and incompatible assumptions about authority.

3.6. Security and Accountability Are Incomplete

Without a model that connects human identity to agent behavior, ecosystems cannot reliably determine whether an agent is acting within its intended authority, whether a delegation chain remains valid, or whether an action can be traced back to a responsible human. This limits the ability of platforms to detect impersonation, prevent unauthorized replication, or enforce accountability across agent interactions.

4. Security and Privacy Considerations

The absence of a consistent model for human-anchored agent identity creates security and privacy risks across ecosystems. These risks arise not from any single technology, but from the structural gaps described in this document. Without a way to connect agent behavior to human authority, systems cannot reliably determine whether an agent is legitimate, whether it is acting within its intended scope, or whether its actions can be traced to a responsible human. Current identity systems do not provide the security or privacy properties needed for autonomous and semi-autonomous agents acting on behalf of people; a consistent architectural model is required to preserve human authority, express delegation, and maintain verifiable provenance while avoiding unnecessary linkability or cross-context correlation.

4.1. Impersonation and Misrepresentation

Agents may impersonate other agents or present themselves as acting on behalf of a human without a verifiable link to that human. Existing identity systems authenticate software components but do not express the human identity root or the delegation chain that authorized the agent. This makes it difficult to detect impersonation or determine whether an agent is acting legitimately.

4.2. Unauthorized Replication

Agents can be copied or instantiated without the knowledge or consent of the human who originally authorized them. Without a model for lineage or controlled replication, ecosystems cannot distinguish authorized instances from unauthorized copies, nor can they determine which instances remain valid over time. This creates opportunities for misuse, fraud, or unbounded agent proliferation.

4.3. Loss of Provenance

As agents act across platforms, the provenance of their actions is often lost. Systems may record that an action was taken by a software component, but not which human authorized the agent, which delegation chain applied, or whether the agent was operating within its intended scope. Without durable provenance, accountability becomes difficult, and misuse may go undetected.

4.4. Privacy Risks

In the absence of a consistent architectural model, ecosystems may rely on ad-hoc identifiers, cross-context correlation, or platform-specific tracking to infer relationships between humans and agents. These practices can erode privacy, create unnecessary linkability, or expose sensitive information about human behavior. A model that preserves human authority must also avoid introducing global identifiers or mechanisms that enable cross-context tracking.

4.5. Fragmentation Across Ecosystems

Because identity systems vary widely, agents may be interpreted differently across platforms, leading to inconsistent assumptions about authority, delegation, and provenance. This fragmentation increases the likelihood of security gaps, misconfigurations, or unintended privilege escalation as agents move across organizational boundaries.

5. IANA Considerations

This document has no IANA actions.

6. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

7. Informative References

[RFC6749] Hardt, D., "The OAuth 2.0 Authorization Framework", RFC 6749, 2012, <https://www.rfc-editor.org/rfc/rfc6749>.

[RFC7643] Hunt, P., Ansari, A., Sanchez, M., and K. McCloghrie, "System for Cross-domain Identity Management: Core Schema", RFC 7643, 2015, <https://www.rfc-editor.org/rfc/rfc7643>.

[DID-Core] Sporny, M., Longley, D., and C. Allen, "Decentralized Identifiers (DIDs) v1.0", W3C Recommendation DID-Core, 2022, <https://www.w3.org/TR/did-core/>.

Author’s Address

Brandon Wesley Beyer
Independent
Email: brandnbyr@icloud.com