

Agent Context Policy Token: DAG Delegation with Human Override

true

2026-02-28

Abstract

This document defines a minimal token profile for safe agent context transfer. The profile has two contributions: (1) an Agent DAG Token for explicit delegation lineage and execution constraints across agents, and (2) a Human-in-the-Loop (HITL) policy mechanism for mandatory human override in safety-relevant situations.

The profile is intentionally narrow. It does not define encryption envelopes, trust onboarding, regulated-environment controls, or broader workflow frameworks. It specifies only interoperable base claims and processing behavior needed for DAG-based delegation and HITL safety intervention.

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.”

Copyright Notice

Copyright (c) 2026 IETF Trust and the persons identified as the document authors. All rights reserved.

1. Introduction

Agentic systems frequently delegate tasks across multiple software agents. Existing ad hoc context passing rarely encodes explicit delegation topology or a standardized safety pause/escalation mechanism. This creates ambiguity in responsibility chains and weakens operational safety.

This specification defines a base token profile that addresses both concerns with minimal overhead:

- **Agent DAG Token:** expresses delegation as a Directed Acyclic Graph (DAG) with explicit node and edge semantics.
- **HITL Policy:** expresses machine-readable conditions under which processing **MUST** be paused for human decision.

The design goal is practical deployment in heterogeneous systems using existing token/signature machinery. Confidentiality and complex governance features are deferred to future extensions.

1.1. Scope

This document defines:

- A minimal claim set for DAG delegation and safety policy.
- Validation and processing rules for interoperable behavior.
- A minimal audit decision record for HITL actions.

This document does not define:

- Encryption envelopes or payload confidentiality mechanisms.
- Regulated-environment control catalogs.
- Trust framework onboarding or remote attestation.
- Full workflow orchestration semantics.

2. Conventions and Terminology

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “NOT RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in BCP 14 {{RFC2119}} {{RFC8174}} when, and only when, they appear in all capitals, as shown here.

2.1. Terms

- **Token Issuer:** Entity that issues the token.
- **Node:** A task or execution step identifier in the delegation graph.
- **Edge:** A directed delegation relation between nodes.
- **Current Node:** Node the processing component is currently executing.
- **HITL Trigger:** A condition requiring human decision.
- **Human Approver:** A human principal authorized to decide override actions.
- **Decision Record:** Minimal audit object capturing human intervention.

3. Safety-Centric Use Cases

3.1. Clinical Triage Assistant

An intake agent delegates symptom summarization to a model agent, then to a recommendation agent. If high-risk symptom keywords are detected, HITL policy requires clinician approval before recommendations are returned.

3.2. Industrial Robot Workcell

A planning agent delegates path optimization to motion agents. If path confidence drops below threshold near human workers, execution pauses and requests supervisor override.

3.3. Financial Fraud Escalation

A fraud scoring agent delegates to investigation agents. If a low-confidence/high-impact action (account freeze) is requested, HITL policy requires analyst confirmation.

4. Requirements

An implementation conforming to this specification:

1. MUST support all required base claims in Section 5.
2. MUST validate DAG acyclicity before acting on delegation context.
3. MUST enforce HITL trigger actions prior to safety-relevant continuation.
4. MUST produce a Decision Record for each HITL decision.
5. MUST fail closed when policy conflicts cannot be deterministically resolved.
6. MUST reject tokens missing required temporal or issuer context.

5. Token Model

This profile is carried as claims within a signed token format (for example, a JWT `{{RFC7519}}`). The exact serialization container is out of scope, provided claim semantics and processing behavior are preserved.

5.1. Base Claims

Required claims:

- `iss` (string): issuer identifier.
- `sub` (string): subject identifier of the principal/process context.
- `aud` (string or array): intended audience.
- `iat` (number): issued-at timestamp.
- `exp` (number): expiration timestamp.
- `jti` (string): unique token identifier.
- `actx_ver` (string): profile version, initially "1.0".

5.2. DAG Claims

Required DAG claims:

- `dag` (object): DAG container.
 - `nodes` (array, REQUIRED): each node object contains:
 - * `id` (string, REQUIRED)
 - * `type` (string, REQUIRED)
 - * `agent` (string, REQUIRED)
 - * `max_depth` (number, OPTIONAL)
 - * `constraints` (object, OPTIONAL)

- **edges** (array, REQUIRED): each edge object contains:
 - * **from** (string, REQUIRED)
 - * **to** (string, REQUIRED)
 - * **purpose** (string, OPTIONAL)
- **root** (string, REQUIRED): root node id.
- **cur** (string, REQUIRED): current node id.
- **path** (array of string, OPTIONAL): observed traversal path.

5.3. HITL Policy Claims

Required HITL claims:

- **hitl** (object): HITL policy container.
 - **version** (string, REQUIRED), initially "1.0".
 - **rules** (array, REQUIRED, at least one item):
 - * **id** (string, REQUIRED)
 - * **trigger** (object, REQUIRED)
 - * **required_role** (string, REQUIRED)
 - * **action** (string, REQUIRED): one of **pause**, **escalate**, **abort**.
 - * **allow_override** (boolean, REQUIRED)
 - * **override_action** (string, OPTIONAL): one of **continue**, **abort**, **reroute**.
 - **unreachable_human** (string, REQUIRED): one of **abort**, **safe_pause**.

5.4. Minimal Trigger Object

A trigger object MUST include:

- **kind** (string): e.g., **risk_score**, **keyword_match**, **confidence_below**.
- **op** (string): one of **gt**, **gte**, **lt**, **lte**, **eq**, **in**.
- **value** (number|string|array): threshold or matcher value.
- **input_ref** (string): processing attribute to evaluate.

6. Processing Semantics

6.1. Validation

Before processing, the receiver MUST:

1. Validate token signature and base temporal checks (`iat`, `exp`).
2. Validate presence and type of required claims.
3. Validate `dag.root`, `cur`, and all edge endpoints reference existing nodes.
4. Validate DAG acyclicity.
5. Validate `cur` is reachable from `root`.

If any check fails, processing MUST stop and return `invalid_token`.

6.2. Delegation and Traversal

When delegating from node A to node B:

1. An edge `A -> B` MUST exist.
2. If `max_depth` applies, resulting depth MUST NOT exceed limit.
3. Node constraints (if present) MUST be enforced before delegation.
4. On success, `cur` MUST be set to B and `path` SHOULD append B.

If delegation check fails, processing MUST stop and return `invalid_delegation`.

6.3. HITL Evaluation

For each safety-relevant operation, receivers MUST evaluate `hitl.rules` in deterministic order by array position.

- If no rule triggers, processing MAY continue.
- If one or more rules trigger:
 - If any triggered rule has `action=abort`, receiver MUST abort.
 - Else receiver MUST apply `pause/escalate` behavior and request human decision per `required_role`.

A human decision MUST result in one of:

- `continue` (if override allowed),
- `abort`,
- `reroute` (if supported and policy allows).

If no human approver is reachable, receiver MUST apply `unreachable_human` behavior.

6.4. Conflict Resolution

If multiple triggered rules prescribe different actions, receiver **MUST** apply strict precedence:

1. **abort**
2. **escalate**
3. **pause**

If ambiguity remains (for example multiple mutually exclusive override actions), receiver **MUST** fail closed with **policy_conflict**.

6.5. Decision Record

Each HITL decision **MUST** produce a Decision Record with at least:

- **decision_id** (string)
- **token_jti** (string)
- **rule_ids** (array of string)
- **human_id** (string)
- **human_role** (string)
- **decision** (string)
- **reason** (string, MAY be empty)
- **time** (number)

7. Interoperability and Deployment Considerations

Implementations **SHOULD** use stable node identifiers for cross-system processing. Unknown optional fields **MUST** be ignored unless explicitly marked critical by implementation policy.

Clock skew handling **SHOULD** be bounded and consistently configured across participants. Deployments **SHOULD** provide deterministic rule ordering and maintain decision logs with integrity protections.

8. Security Considerations

This profile targets safety and control-plane clarity, not confidentiality. Safety risks include bypassed HITL checks, forged delegation paths, replay of stale tokens, and policy downgrades.

Implementations **SHOULD**:

- enforce token signature verification and expiration,
- bind issuer trust to deployment policy,
- detect replay using `jti` where feasible,
- prohibit silent fallback when HITL evaluation fails,
- minimize override authority scope by role.

Confidentiality is explicitly out of scope for this base profile. Encrypted envelope mechanisms MAY be defined in future extensions.

9. Privacy Considerations

Tokens can reveal workflow topology and role metadata. Implementations SHOULD avoid unnecessary personal data in node, rule, and decision claims.

Decision Records SHOULD use least-identifying human identifiers consistent with accountability requirements, and retention windows SHOULD be minimized per deployment policy.

10. IANA Considerations

This document has no IANA actions.

11. References

11.1. Normative References

- {{RFC2119}}
- {{RFC8174}}
- {{RFC7519}}

11.2. Informative References

- {{ECT}}
- {{TOKEN-CONTAINER}}

Appendix A. Compact JSON Examples

A.1. Example DAG Token Claims

```
{
  "iss": "https://issuer.example",
  "sub": "workflow:triage-42",
  "aud": "https://runtime.example",
  "iat": 1771939200,
  "exp": 1771942800,
  "jti": "9b524a7c-f2b8-4f41-9f23-472f63f24c95",
  "actx_ver": "1.0",
  "dag": {
    "root": "n0",
    "nodes": [
      { "id": "n0", "type": "intake", "agent": "agent:intake" },
      { "id": "n1", "type": "summarize", "agent": "agent:llm", "max_depth": 3 },
      { "id": "n2", "type": "recommend", "agent": "agent:reco" }
    ],
    "edges": [
      { "from": "n0", "to": "n1", "purpose": "summarization" },
      { "from": "n1", "to": "n2", "purpose": "recommendation" }
    ]
  },
  "cur": "n1",
  "path": ["n0", "n1"]
}
```

A.2. Example HITL Policy Claims

```
{
  "hitl": {
    "version": "1.0",
    "unreachable_human": "safe_pause",
    "rules": [
      {
        "id": "r-high-risk",
        "trigger": {
          "kind": "risk_score",
          "op": "gte",
          "value": 0.85,
          "input_ref": "eval.risk"
        },
        "required_role": "clinician:oncall",
      }
    ]
  }
}
```

```

        "action": "escalate",
        "allow_override": true,
        "override_action": "continue"
    },
    {
        "id": "r-low-confidence",
        "trigger": {
            "kind": "confidence_below",
            "op": "lt",
            "value": 0.6,
            "input_ref": "eval.confidence"
        },
        "required_role": "clinician:oncall",
        "action": "pause",
        "allow_override": true,
        "override_action": "reroute"
    }
]
}

```

A.3. End-to-End Processing Example

```

{
  "event": "hitl_decision",
  "decision_id": "dec-2f5a9f77",
  "token_jti": "9b524a7c-f2b8-4f41-9f23-472f63f24c95",
  "rule_ids": ["r-high-risk"],
  "human_id": "user:alice",
  "human_role": "clinician:oncall",
  "decision": "continue",
  "reason": "reviewed chart context",
  "time": 1771940102
}

```

Delta Note: Intentionally Deferred in This Base Draft

This base draft intentionally excludes encryption envelopes, regulated controls, remote attestation, trust onboarding, and broad workflow orchestration to keep interoperability achievable for early implementations.

Why these are deferred

- **Encryption envelope:** confidentiality mechanisms add key management, envelope formats, and processing complexity that are orthogonal to core delegation and safety semantics.
- **Regulated controls:** jurisdiction-specific controls differ significantly and can be layered as profiles once base behavior is stable.
- **Remote attestation:** useful for stronger trust signals but not required to establish token-level delegation/HITL interoperability.
- **Trust onboarding:** federation metadata and legal trust framework artifacts are deployment concerns outside minimal protocol semantics.
- **Rich orchestration semantics:** full workflow languages and compensation logic exceed the scope of a compact token profile.

Extension path

Future extensions can define:

1. Confidential profile (encrypted payload and selective disclosure).
2. Regulated profiles (health, finance, public sector control sets).
3. Trust profile (attestation claims and verifier behavior).
4. Advanced policy profile (multi-party approvals, quorum, time-bounded overrides).

This sequencing keeps the base draft focused on the immediate gap: explicit DAG delegation context plus deterministic human safety override behavior.