Human Seal

OPEN STANDARD · FREE TO IMPLEMENT · VENDOR-NEUTRAL

The Human Seal Standard

An open standard for proving that a real, identifiable human authorized a world-affecting action taken by an AI system. It defines what a valid human approval is, how it is recorded, and how any party verifies it offline. Free to implement. Belongs to no single vendor.

V0.1 PUBLIC REVIEW DRAFT

PUBLISHED BY	DATE	TRUST ROOT	REFERENCE IMPL
Human Seal (humanseal.world)	18 September 2026	did:web:humanseal.world	FLOCORE

1 · PURPOSE AND SCOPE

1 Purpose and scope

The Human Gate is the control point at which a world-affecting action proposed by an AI system is held until a real human approves it. This specification defines that control point so it means the same thing everywhere: what counts as a valid human approval, how the approval is recorded, and how any counterparty can verify it without trusting the system that produced it.

This standard defines two things: the Human Gate, the checkpoint where a real human approves an AI world-affecting action, and the Human Seal, the signed, verifiable proof that they did. It is open and free to implement, in the model of Bluetooth or the web. Human Seal stewards the definition and the mark; it does not sell the gate. The keywords MUST, MUST NOT, SHOULD and MAY are used as in RFC 2119.

2 · DEFINITIONS

2 Definitions (normative)

- **World-affecting action:** any action by an AI system that changes state outside itself: a payment or transfer, a write to a system of record, a message sent on someone's behalf, a deletion, a booking, an order, a configuration change.
- **Human Gate:** the control point at which a world-affecting action is held until a human approves it.
- **Approver:** a real, identifiable human, distinct from the AI agent, authenticated to a level appropriate to the action.
- **Assent:** the formal act by which the Approver agrees to a specific action at the gate. Assent is per action, informed, and never a standing or blanket consent.
- **Receipt:** a signed, tamper-evident record of the Assent (or refusal): that a specific action was approved or refused by a specific Approver at a specific time.
- **Issuer:** the party that signs Receipts, identified by a domain-anchored trust root (did:web).

3 · NORMATIVE REQUIREMENTS

3 Normative requirements

HG-1	A conforming system **MUST NOT** execute a world-affecting action without an approval Receipt for that action.
HG-2	The Approver **MUST** be a human identity distinct from the agent that proposed the action.
HG-3	The approval **MUST** be bound to the specific action (its action id and a content hash of its payload), not a blanket or standing consent.
HG-4	Every approval and every refusal **MUST** produce a Receipt, recorded in an append-only log.
HG-5	The Receipt **MUST** be cryptographically signed by the Issuer and verifiable offline against a published key.
HG-6	The Issuer key **MUST** be resolvable via a stable, domain-anchored trust root (did:web).
HG-7	The authentication strength of the Approver **SHOULD** scale with the risk of the action (a step-up for money or access).
HG-8	A conforming system **MUST NOT** provide an autonomous path that bypasses the gate for a world-affecting action.

4 · THE RECEIPT

4 The Receipt (schema)

A Receipt is a JSON object. Personal data of the Approver is referenced by a stable identifier, never embedded.

```
{
  "action_id":     "string, unique per action",
  "action_type":   "payment | write | message | delete | booking | config | ...",
  "content_hash":  "sha256 of the canonical action payload",
  "tenant":        "the accountable organisation",
  "approver_ref":  "stable id of the human approver (not PII)",
  "decision":      "approved | refused",
  "decided_at":    "RFC3339 timestamp, UTC",
  "issuer":        "did:web:humanseal.world",
  "signature":     "ed25519 over the canonical receipt",
  "public_key":    "base64url raw 32-byte Ed25519 key"
}
```

Canonicalization: the object is serialized with sorted keys and no insignificant whitespace before signing, so any verifier reproduces the exact bytes.

5 · VERIFICATION

5 Verification (offline)

A counterparty verifies a Receipt without contacting the Issuer:

- Resolve the **issuer** did:web over HTTPS to obtain the Ed25519 public key (anchored by DNS and TLS to the Issuer's domain).
- Recompute the **canonical** form of the Receipt.
- Check **signature** over the canonical form against the public key.
- Check that **content_hash** matches the action being relied upon.

If all four hold, the approval is authentic and non-repudiable. No call to the Issuer is required, so a Receipt proves itself even years later.

6 · CONFORMANCE

6 Conformance levels

LEVEL	NAME	REQUIRES
1	Gated	HG-1 to HG-4. Actions are gated by a human and logged.
2	Signed	Level 1 plus HG-5, HG-6. Receipts are signed and offline-verifiable.
3	Assured	Level 2 plus HG-7, HG-8 and a present-human proof (anti-mimicry step-up).

The **Human Seal** is awarded to systems that pass the conformance suite at a stated level (for example "Human Seal, Level 2"). The registry of sealed systems is public at humanseal.world. The reference implementation (FLOCORE) publishes a live did:web trust root and Ed25519-signed receipts today, demonstrating Level 2 is implementable now.

ANNEX A

A EU AI Act: logging and human oversight

Two articles of the EU AI Act read almost like a specification for the Human Gate. **Article 12** requires high-risk AI systems to allow the automatic recording of events (logs) over the system's lifetime, with traceability appropriate to the intended purpose. The Human Seal satisfies it as follows.

ARTICLE 12 OBLIGATION	SATISFIED BY
Automatic recording of events over the lifecycle	HG-4: every approval and refusal produces a Receipt in an append-only log.
Traceability of each event to the action and actor	HG-3: the Receipt binds the decision to a specific action id and content hash; approver_ref identifies who decided.
Identification of situations that may present a risk	Refusals and withheld actions are logged as first-class events, surfacing the risky situations rather than hiding them.
Integrity and non-repudiation of the logs	HG-5: signed, tamper-evident Receipts make each log entry verifiable and non-repudiable offline.
Availability over the retention period	The append-only log is durable, and each Receipt is self-verifying independent of the Issuer's continued operation.

Article 14 requires human oversight of high-risk AI: the ability to understand, oversee, intervene in, override and stop the system, with a two-person check for certain biometric uses. This is the human-approval requirement itself, and it is what the Human Gate is.

ARTICLE 14 OBLIGATION	SATISFIED BY
A human can intervene in and override the system	HG-1: no world-affecting action executes without a human approval; the action is held, not run.
Oversight by a natural person distinct from the system	HG-2: the Approver is a human identity distinct from the agent that proposed the action.
A stop or reject capability	Every action can be refused at the gate, and the refusal is itself a logged, signed event.
Guarding against automation bias / over-reliance	HG-8: no autonomous path may bypass the gate; HG-7: authentication steps up with risk.
Two-person verification for certain high-risk uses	Level 3 (Assured): a present-human, risk-scaled step-up supports a second-approver check.

Mapping for legal review under the EU AI Act. Not legal advice. Conformance supports, but does not by itself establish, regulatory compliance, which depends on the full system context and qualified counsel. The Article 12 and 14 texts are summarized; the high-risk timeline rests on the provisional 2026 arrangement.

ANNEX B

B South Africa: POPIA Section 71

POPIA Section 71 gives a data subject the right not to be subject to a decision based solely on the automated processing of their information where it has a legal or material effect, and requires that a responsible party provide an opportunity for human intervention. This is live South African law today, and the Human Gate addresses it directly.

POPIA S71 OBLIGATION	SATISFIED BY
A decision may not be based solely on automated processing	HG-1 and HG-2: a real human approves the world-affecting decision, so it is no longer solely automated.
Provide the opportunity for human intervention	The gate is exactly that opportunity: the action waits for a person to assent or refuse.
Evidence that the human step occurred	HG-4 and HG-5: the signed Receipt (the Human Seal) is durable proof that a specific person intervened, and when.

Mapping for legal review under POPIA. Not legal advice. Together, Annex A and Annex B show that one artifact, the signed Human Seal receipt, evidences EU Article 12 logging, EU Article 14 human oversight, and POPIA s71 human intervention at once.

STATUS

7 Status of this document

This is version 0.1, a Public Review Draft published by Human Seal for comment. The trust root of the standard is **did:web:humanseal.world**, the registered domain that publishes the Ed25519 key every Human Seal is verified against. The reference implementation is FLOCORE, which issues and signs Seals against that root; the did:web document goes live the moment humanseal.world DNS points to the reference server. Legal and regulatory statements are for review by qualified counsel and are not legal advice.