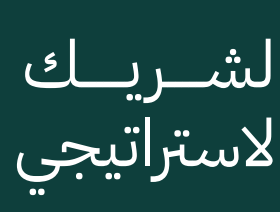
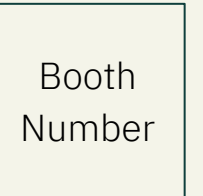




UP2U2B LLC
Riyadh · Saudi Arabia
أب2بي



TAWTHIQ — Tamper-Evident Provenance Chain for Human Decisions over AI



Student Name: Emad Suleiman Alwan

ID:

Introduction

AI now drafts analyses, reports and recommendations inside security institutions. After any incident, investigators ask the same questions: **who approved this AI output, when, on what basis — and has the record changed since?**

Ordinary logs cannot answer with certainty: an administrator can edit them, and cloud audit trails require trusting the vendor.

TAWTHIQ (توثيق, “documentation”) turns every human decision over an AI output into forensic evidence that a third party can verify offline, years later, **without trusting us**.

Who approved?

When?

Changed since?

Three questions every investigation asks — answered by evidence, not by trust.

Methodology

1 Hash & link

Each record is hashed with SHA-256. Its seal commits to the previous seal and to nine canonical fields (content hash, witness, intent, stage, Hijri timestamp...). Changing one byte breaks the chain at that record.

3 External time witness

Chain-head digests are time-stamped with OpenTimestamps and anchored in Bitcoin — outside our systems.

2 Human signature

The owner signs records with a hardware-bound key (OpenSSH signatures, dedicated namespace). The chain head is sealed and signed periodically.

4 Independent verifier

One exported folder + a standard-library Python script + ssh-keygen: any auditor re-checks every link and signature and compares each signed hash with the ledger — on any computer, with none of our code.

Principles: local-first — records never leave the machine, only 32-byte digests go to the time-stamp calendars · Arabic-first governance — the *intent* of each record is an indexed field · acceptance criteria are frozen before results are read.

Results

Measured on **27 Sep 2026** by the independent verifier on the live chain — and **sealed in the chain** as record #2,128 (tawthiq-20260927T020024817751), signed by the owner:

2,127

records sealed since 4 June 2026

2,126

signatures, all valid (2,097 hardware-key)

73

chain seals, all valid

32

OpenTimestamps stamps (2 confirmed, block 965242)

5 broken links detected and named — caused by our own ledger merge (records of 7–9 Sep 2026). The verifier caught its builder's error; nothing was hidden or rewritten.

Also: 2,078 seals recomputed and matching the contract; 44 superseded drafts declared. Limits: single writer, not distributed; key residency in hardware is asserted by setup, not provable from the public key.

Innovation

Hashing, signatures and time-stamping are each established. The innovation is their **composition in live use** for a question AI governance cannot yet answer: *can a human's decision over an AI output be proven later, by anyone?*

- Provenance of the **human decision**, not only the file
- **Intent** recorded with the act — why, not only what
- **Offline, dependency-free** verification — no trust in the provider
- **Arabic-first**, sovereign, local-first deployment
- **Measured self-correction** — faults are published, not patched silently

Security applications: chain of custody for digital evidence · non-repudiation of approvals · audit of AI-assisted decisions.

Conclusion

Most AI-governance tools *describe* compliance. **TAWTHIQ lets anyone check it:** who approved what, and when — on any computer, without our code and without trusting us.

Tamper tests on copies of the live chain: a one-byte edit → **caught**; a forged record with every later seal recomputed → **caught** by the signed seals.

Actual verifier output · 27 Sep 2026 · Arabic-first tool

■ الدفتر: 2127 حلقة · الرأس 3efabf9ca4c45286 · كسوف الربط: 5
■ ختوم السلسلة: 73 · صحيح التوقيع 73
■ توافيق الحلقات: 2126 · صحيح 2126

Future Work & References

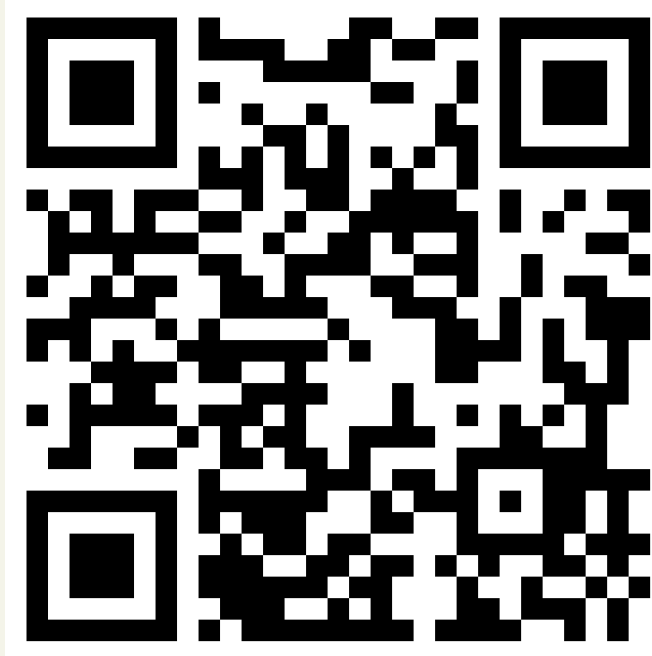
Future work

- Typed decision records (model-output & prompt hashes, declared model, human action), enforced by the verifier; results rejected without a prior frozen criterion
- Fork-safe merge that re-links branches without rewriting history
- Pilot with a security unit on non-classified data, criteria fixed in advance

References

1. Haber & Stornetta, “How to time-stamp a digital document”, *J. Cryptology* 3(2), 1991
2. NIST FIPS 180-4, Secure Hash Standard (SHA-256)
3. OpenSSH, PROTOCOL.sshsig (ssh-keygen -Y)
4. OpenTimestamps — opentimestamps.org

IP: copyright reg. 26-12-95029589 (SAIP) · trademark application SA-1268915



Supporting files & verification bundle
up2u2b.com/tawthiq/