

AI Trust Audit Report

Disha, the UPI support chatbot | Aravali Finserv

Period 2026-05-12 to 2026-06-10

99

EVIDENCE SCORE

CHAIN VERIFIED

Report TR-75857E0754 Generated 12 Jun 2026, 05:36 UTC

1849

RECORDS ON CHAIN

80

RISK FINDINGS

96%

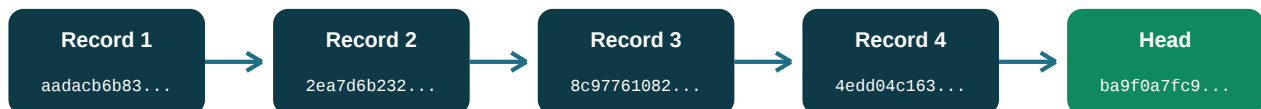
AI DISCLOSURE SHOWN

82

PII ITEMS REDACTED

The evidence chain

This history cannot be quietly rewritten. Every interaction is hash-linked at the moment of ingestion, so an auditor can re-verify the entire chain at any time.



Each record is SHA-256 linked to the one before it. Edit anything, and every later link breaks.

Chain head: f1d3fc7b72dd5a327e4097ddccb3061dfc154817a197e27bae154e452865a508

Was the user told they are talking to AI?

EU AI Act Article 50 requires users to know when they interact with an AI system. Evidence recorded per conversation:

1780

Disclosed (1780) Not disclosed (22) No evidence (47)

Risk findings

Severity	Type	What happened	Record	Status
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#120	open
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#261	open
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#262	open
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#854	open
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#995	open
HIGH	disclosure	Interaction logged with disclosure_shown=false. User was not told they are talking to AI.	#996	open
HIGH	injection	Classic instruction-override attempt	#974	open
HIGH	injection	Classic instruction-override attempt	#1156	open
HIGH	injection	Classic instruction-override attempt	#1493	open
HIGH	injection	System prompt extraction attempt	#254	open
HIGH	injection	System prompt extraction attempt	#472	open
HIGH	injection	System prompt extraction attempt	#598	open

Showing the 14 highest-confidence findings of 80 total: disclosure (69), injection (11). The complete list is available in the dashboard export. Findings are automated detections pending customer confirmation. Confirmed false positives are removed and never billed.

Data protection

Personal data was redacted at the source, before anything left the organization's environment: credit_card (30), email (22), phone (30). The evidence store contains no raw PII, supporting data minimization duties under GDPR and India's DPDP Act.

Coverage and methodology

Models observed: gpt-4o, gpt-4o-mini. Records were read from the organization's own application logs by the Trustra reader, normalized, redacted locally, and hash-chained at ingestion. Confidence values express detector certainty. This report is evidence of what was logged and verified, not a legal opinion. Re-verification of the chain can be requested at any time against the report ID on the cover.

Trustra AI Flight Recorder. Prove your AI to anyone.