The evidence standard for claims of control

Document AIRA-PI-2026-0104 · Published 2 September 2026 · Issuing office: Public Information Office

Human responsibility for AI systems

Who controls AI agents? And who remains responsible when an agent acts across sensitive information and human decisions?

Operators remain responsible for AI activity that affects access, information or decisions. They should maintain records that allow this activity to be reviewed.

A.I.R.A. publishes operator guidance, receives reports of concerns and makes incident findings available to the public.

- Who can stop it?
- Who can inspect it?
- Who answers for it?

The evidence standard for human control of AI

Published findings should identify their sources, review dates and correction history. Operators should be able to demonstrate who is authorised to intervene and how system activity can be inspected.

Attributable sources

Findings identify the evidence used in review.

Visible review dates

Records show when they were last reviewed, not just when they were written.

Correction history

Changes stay beside the original entry, with the reason logged.

Proven intervention

Records document a person’s authority and ability to pause the system.

The public interest

AI systems can support research and public services. Oversight helps people understand their use, raise concerns and review how incidents are handled.

Fictional record created for the Who Controls AI Agents? campaign and CTRL: Locust. It holds no legal authority.

Responsibility remains with the named operator throughout system operation and incident review.

Fictional record created for the Who Controls AI Agents? campaign and CTRL: Locust. It holds no legal authority.