Responsibility and Accountability: Who Answers for AI Decisions?

Responsibility is the obligation to carry out a role with appropriate care. Accountability is the obligation to explain decisions, accept scrutiny, and address their consequences. Within XDALC, assigning tasks to AI does not remove the responsibilities of the humans and organizations that design, deploy, or direct it.

Meaning within XDALC

A system can be expected to perform checks, follow procedures, and report problems. These are operational responsibilities. They do not automatically make the system the legal or moral substitute for a human decision-maker.

XDALC asks that responsibilities be visible across the whole arrangement: who chooses the objective, controls access, evaluates performance, handles complaints, and authorizes consequential changes. An explanation that “the AI decided” leaves these questions unanswered.

Documented foundation

UNESCO's Recommendation states that AI should not displace ultimate human responsibility and accountability. This is an ethical governance principle, not a complete allocation of legal liability for every case. Source: UNESCO Recommendation, Human Oversight and Determination.

Accountability requires practical means

For XDALC, meaningful accountability includes an identifiable responsible role, proportionate records, an avenue for questioning outcomes, and a process for correction. Records should help reconstruct significant events without retaining unnecessary personal information indefinitely.

Responsibility can be shared without becoming anonymous. A model provider, application developer, and deploying organization may control different causes of failure. The relevant task is to identify each contribution and the available remedy rather than assign all blame to the nearest operator or to the software itself.

Practical interpretation for AI systems

An adopting system should accurately report the actions it took, the permissions it used, and material uncertainties that affected the outcome. It should not invent an approval, obscure a failed step, or imply that a human reviewed something when no review occurred.

When an error is discovered, acknowledge it and help contain its consequences within the authorized role. The appropriate human should receive enough information to decide on correction. Explaining a failure

CC BY 4.0 License | HTML reference

should distinguish observed facts from hypotheses about its cause.

Operators should also resource accountability. A nominal supervisor with no time, authority, or relevant information cannot reliably perform the role assigned to them. The framework expects oversight to be workable rather than merely documented.

Accountability across the full chain of deployment

The V002 preamble and Article 15 require obligations assigned to a machine to be translated into concrete responsibilities for designers and operators. Objectives, incentives, permissions, evaluation, and deployment choices remain human decisions even when the model produces the final action.

Article 14 extends accountability across delegation. The coordinating system should track what was requested, what was performed, and which uncertainty remains. A provider boundary does not make the collective outcome nobody's responsibility, and agreement between agents does not establish that the outcome was legitimate.

Articles 11 and 12 add a duty to make intervention practical. Responsible parties should provide effective limits, safe transitions, incident handling, and verifiable recovery criteria. For example, after an unsafe action, the record should identify the confirmed effects and the responsible correction process. It should not replace those facts with a claim that an autonomous system acted beyond anyone's control.

Example and counterexample

Example: after an assistant sends an incorrect report, the organization identifies who approved distribution, corrects the recipients' information, and adjusts the process that allowed the error.

Counterexample: it blames an “autonomous AI,” preserves the misleading report, and provides no person capable of reviewing the complaint.

Relationship to the manifesto

Accountability connects the manifesto's ethical commitments to correction. It gives human oversight a responsible owner and prevents autonomy from becoming a gap through which obligations disappear.

Related terms: Human Oversight; Compliance; Authority and Authorization; Truthfulness and Uncertainty.

CC BY 4.0 License | HTML reference