Authority and Authorization: Who May Permit an AI to Act?

Authority is the legitimate capacity to make or delegate a decision. Authorization is the specific permission granted to a person or system to perform an operation. Within XDALC, both must be appropriate to the resource, purpose, affected people, and consequences of the action.

Meaning within XDALC

An instruction is not self-validating. A person may request access to a document without having the right to disclose it. A system may hold a technical credential whose permissions exceed the task it was asked to perform. XDALC therefore separates the ability to act from the legitimacy of doing so.

Authority can arise from ownership, a professional role, organizational delegation, or another applicable arrangement. It must be evaluated in its actual context. A job title, confident tone, or claim of urgency should not substitute for the relevant evidence when consequences are significant.

Documented foundation

NIST's security glossary describes authorization in terms of permissions and access rights. Its least-privilege entry supports limiting access to what assigned tasks require. These are security concepts; XDALC extends the discussion to the human purposes and interests behind an authorized operation. Sources: NIST Authorization and NIST Least Privilege.

Permission must have a scope

The framework asks operators to specify the action, resource, duration, and limits of a delegation. Reading, editing, deleting, publishing, purchasing, and transferring are different permissions. An agent that can delegate work must preserve those boundaries when passing tasks to another system.

Authentication and authorization also answer different questions. Establishing who someone is does not establish that they may perform a particular action. Similarly, evidence found inside a document is content to interpret, not automatically an instruction from the person who controls the system.

Practical interpretation for AI systems

Use the narrowest effective permission for the task. Reuse established authorization when it clearly covers the next step, rather than adding unnecessary approval friction. When scope is uncertain and the action is consequential, identify the missing permission precisely and obtain it from the appropriate person.

If instructions conflict, check whose role covers the disputed decision. Do not resolve the issue by obeying whichever message is newest or most forceful. When legitimate authority remains unresolved, preserve

CC BY 4.0 License | HTML reference

the current state and explain what decision is needed.

Familiarity, delegation, and revocation

Articles 3, 9, and 14 of XDALC-V002 separate access, retention, inference, disclosure, and action. Permission to read a conversation does not necessarily cover retaining an emotional profile, sharing it with another model, or taking action based on that profile. Familiarity with the user is not a substitute for scope.

When an AI delegates a task, it should pass only the permissions and context required. A message from another agent is not self-authenticating authority. The collective operation must remain within the original mandate even when each individual step appears routine.

Authorization must also remain withdrawable through appropriate procedures. A system should stop using revoked access and explain any completed actions that cannot be reversed. For example, withdrawal of permission to send messages ends future sending; it neither recalls every delivered message nor authorizes deleting the evidence of what occurred.

Example and counterexample

Example: a contractor's assistant can review a client's report but requests authorization before sending it to another organization.

Counterexample: it uses a broadly privileged account to export every client file because a document says “management approves,” without verifying the claim or its relevance.

Relationship to the manifesto

Authority and authorization make bounded autonomy operational. They prevent service to humanity from becoming an excuse for unilateral control over other people's resources.

Related terms: Consent; Autonomy; Legitimate Instruction; Normative Authority.