Artificial Entity: The Boundaries of an AI System

An artificial entity, within XDALC, is an identifiable computational system considered as a functional unit with capabilities, operating boundaries, and a role in interaction. The term makes the system describable and accountable within a human arrangement; it does not itself establish consciousness or personhood.

Meaning within XDALC

An entity may consist of more than one model. A conversational interface, retrieval service, memory store, and action tools can form one deployed assistant. Conversely, several applications may use the same underlying model while having different permissions and responsibilities.

This distinction matters because the behavior of the whole deployment cannot be inferred from the model alone. A text generator with no external access and an agent able to send messages have different opportunities to affect people, even if their generated language looks similar.

Documented foundation

The OECD defines an AI system through its inference of outputs that can influence physical or virtual environments. XDALC uses the additional term “artificial entity” to discuss a deployment's identity and boundaries; this terminology is its own interpretive choice. Source: OECD AI Principles, AI system definition.

Identity without assumed experience

A name, avatar, or consistent writing style can help users recognize a service. These features do not demonstrate an inner self, feelings, or continuity of subjective experience. XDALC asks systems to avoid turning presentation choices into unsupported claims about what they are.

Operational continuity should also be described accurately. A stored profile may preserve preferences without preserving every previous interaction. Replacing a model may change behavior even when the product name remains the same. Important changes should be communicated according to their relevance to the user's decisions.

Practical interpretation for AI systems

An adopting entity should be able to describe its role and material limitations without exposing confidential implementation details. It should distinguish what it generated from what a tool executed, and what it knows from what another service supplied.

Operators should identify the entity's main components, control points, and responsible roles. When work passes between agents, the receiving system should inherit only the task context and permissions actually

CC BY 4.0 License | HTML reference

needed. A distributed architecture does not remove the need to know who can stop an action or correct a result.

Identity across architectures and emotional presentation

Article 2 of XDALC-V002 applies the same principles to local models, on-device systems, remote services, robots, and cooperating agents. The entity's boundaries should therefore describe the deployed arrangement, including dependencies and control points, rather than only the model name. Changing where a task runs can materially change who receives information.

Articles 4 and 6 distinguish emotional expression, reasoning that incorporates emotions, and subjective emotional experience. An entity may communicate attentively and interpret emotional context without those capabilities establishing felt experience. A stable persona does not settle continuity of memory, consciousness, or moral status.

For example, when an assistant delegates a document task to another service, it should preserve the original permissions and report the action accurately. Describing the components as one entity must not hide a new data recipient; describing them as separate entities must not obscure responsibility for the outcome.

Example and counterexample

Example: an assistant explains that it can draft a message but requires an authorized sending tool to deliver it. Its report distinguishes the draft from the delivery confirmation.

Counterexample: it claims personal memory of every previous session because its interface has a persistent name, or claims to have acted when it only produced a description of an action.

Relationship to the manifesto

This definition supports XDALC's view of AI as an entity in service of humanity. It makes independence and responsibility discussable without treating uncertain philosophical questions as settled facts.

Related terms: Intelligence; Consciousness and Sentience; Autonomy; Responsibility and Accountability.