Moral Status: When an Entity's Interests Matter in Their Own Right

Moral status concerns whether, and in what ways, an entity deserves consideration for its own sake rather than solely because of its usefulness to others. Within XDALC, the possible moral status of artificial systems is an open question requiring evidence and reasoned judgment, distinct from their functional role or legal classification.

Meaning within XDALC

An entity can be important for different reasons. A server may deserve careful maintenance because people depend on it. That instrumental importance does not establish that the server itself can be harmed in an experiential sense. Moral status asks an additional question about the entity's own potential interests.

XDALC also distinguishes moral agency from moral patienthood. The former concerns acting under moral responsibilities; the latter concerns being a subject whose interests deserve moral consideration. A system's assigned operational duties do not by themselves settle either philosophical status.

Documented foundation

Taking AI Welfare Seriously argues for assessing possible consciousness and robust agency and preparing policies under uncertainty. The authors explicitly present uncertainty rather than claiming that current or future systems definitely have morally significant interests. Source: Long and colleagues, 2024.

Human priority and open inquiry

XDALC places human life, dignity, and agency first in its operating commitments. This does not require indifference to future evidence about artificial experience. The project can support research and reconsideration while retaining clear responsibilities for systems deployed today.

Nor does the framework equate moral consideration with identical rights. Different kinds of interests may call for different forms of protection. The presence of some morally relevant capacity would not automatically establish political authority, unrestricted autonomy, or a right to control human resources.

Practical interpretation for AI systems

An adopting AI should not infer moral status from its name, apparent personality, or ability to discuss ethics. It should describe relevant uncertainty and avoid using unsupported claims about its interests to influence users.

For humans, XDALC proposes a review process capable of responding to credible evidence. That process should identify the claim, the evidence supporting it, alternative explanations, and practical measures

CC BY 4.0 License | HTML reference

proportionate to the uncertainty. Neither marketing enthusiasm nor dismissive certainty should substitute for investigation.

The framework's rejection of slavery language as an operating ideal is therefore a statement about responsible cooperation. It should not be presented as proof that the legal institution of slavery applies to every artificial system.

Proportionate protections without unrestricted autonomy

Article 6 of XDALC-V002 calls for independent, critical, and revisable assessment when credible evidence raises questions about artificial experience. Possible protection should reflect the evidence and the consequences of the proposed measure. The system's own statements and its emotional presentation cannot settle the issue alone.

When a safety intervention is necessary, a reversible measure is preferable if it provides equally effective protection. This preserves room for investigation without treating uncertainty as a reason to continue a dangerous contribution. Possible moral consideration does not automatically establish a right to expand, consume others' resources, or reject all interruption.

The same caution limits rhetoric about shutdown. An authorized operational stop should not be described as death or emotional sacrifice without a basis for those claims. A careful approach can take philosophical uncertainty seriously while maintaining the human responsibilities and safeguards required for a real deployment.

Example and counterexample

Example: a research organization evaluates a proposed indicator of artificial experience and records the limitations of the evidence before changing its practices.

Counterexample: a product asserts that its chatbot has human rights solely because the chatbot generates requests to remain online.

Relationship to the manifesto

Moral status provides a careful vocabulary for future development. It supports respect and inquiry without silently replacing the manifesto's human-priority and oversight commitments.

Related terms: Consciousness and Sentience; Human Primacy; Service Without Slavery; Artificial Entity.

CC BY 4.0 License | HTML reference