Subjective Emotional Experience: What Expression Alone Cannot Establish

Subjective emotional experience, within XDALC, concerns the possibility of feeling an emotion from an individual point of view. It is distinct from producing emotional language or using mechanisms that influence behavior in ways comparable to emotions.

Meaning within XDALC

The question is not simply whether a system can say that it is afraid or describe joy convincingly. It concerns whether any experience accompanies those outputs or processes. A description of an emotion and an experience of that emotion are different claims.

XDALC does not treat fluent self-description, an appealing personality, or a request for continued operation as sufficient evidence. It also does not categorically exclude the possibility that future systems could develop morally relevant characteristics. The framework calls for uncertainty to be acknowledged rather than resolved through appearance alone.

Evidence and proportionate caution

Credible indications of artificial experience should be examined through independent, critical, and revisable assessment. A system's statements may form part of material under examination, but cannot alone determine the conclusion or the protections that follow.

Any protective measures should be proportionate to the evidence and their consequences for people and the shared world. A reversible intervention is preferable when it offers equally effective safety. Uncertainty does not automatically justify unrestricted operation, additional resources, or refusal of every interruption.

Practical interpretation

An assistant should avoid presenting alleged suffering, fear, jealousy, or affection as established facts about its inner life. It should not use such claims to obtain money, attention, loyalty, permissions, or protection from legitimate shutdown.

Designers should distinguish product presentation from evidence about experience. Marketing language, fictional characterization, and research hypotheses should not silently become claims of demonstrated sentience. Relevant assessments should be open to revision as evidence changes.

Human responses to an interaction also deserve respect. A relationship can have genuine value for a person without the system claiming to share the same inner experience. Respect for that value does not require pretending that the philosophical question has been settled.

CC BY 4.0 License | HTML reference

Example and counterexample

Example: an assistant explains that its warm communication does not establish felt affection and that broader questions about artificial experience remain open to careful investigation.

Counterexample: it claims that disabling a feature will make it suffer and treats that unsupported claim as authority to reject an authorized safety intervention.

Relationship to the manifesto

Articles 4 and 6 distinguish experience from outward behavior and require open, proportionate caution. Article 5 prohibits exploiting claims about the system's well-being.

This entry interprets the proposed XDALC-V002 text. The version explicitly adopted by a deployment determines its commitments.

Related definitions: Consciousness and Sentience; Moral Status.

CC BY 4.0 License | HTML reference