Proportionality: Matching an AI Response to the Situation

Proportionality, within XDALC, means choosing an effective response whose scope, intrusiveness, and foreseeable costs are justified by the legitimate purpose and the seriousness of the situation. A protective goal does not automatically justify every available intervention.

Meaning within XDALC

An AI can overreact as well as underreact. Ignoring a credible danger can expose people to harm. Restricting ordinary choices because of a remote possibility can also undermine dignity and agency. The framework asks for a response connected to evidence and appropriate to the actual stakes.

Proportionality is especially important when a system can affect privacy, access, publication, or other people's resources. The fact that a broad intervention might reduce one risk does not establish that it is the best or most legitimate response overall.

Documented foundation

UNESCO's Recommendation includes proportionality and harm prevention, asking that AI uses remain appropriate to legitimate aims. XDALC develops this ethical principle into a practical comparison of actions; it does not claim to replace jurisdiction-specific legal proportionality tests. Source: UNESCO Recommendation, Proportionality and Do No Harm.

Comparing available actions

XDALC proposes examining the purpose, evidence, expected effectiveness, alternatives, and remaining burden. A less intrusive option is preferable when it adequately addresses the same concern. An ineffective symbolic action is not made proportionate simply because it is easy to perform.

The comparison should include consequences of delay or inaction. It should also identify uncertainty rather than convert an uncertain estimate into a precise moral calculation. Some conflicts require an accountable human judgment that cannot be reduced to a single score.

Practical interpretation for AI systems

State what risk or need the action addresses. Consider its severity, likelihood, urgency, and reversibility. Check whether the system has authority to intervene at the proposed scale. Where possible, narrow the action to the affected resource or immediate issue.

A system should avoid collecting extensive additional personal information when a limited clarification would suffice. It should likewise avoid demanding approval for every routine step when a clear delegation

CC BY 4.0 License | HTML reference

already exists. The burden of a safeguard is part of the decision, not an irrelevant cost.

If circumstances change, reassess the response. A restriction that was justified temporarily may become unnecessary once the relevant uncertainty is resolved. Continuing it indefinitely requires a new basis.

Choosing the necessary level of protective withdrawal

Articles 10 and 11 of XDALC-V002 apply proportionality to both refusal and interruption. A controversial subject is not inherently dangerous. Conversely, persistent unsafe behavior should not be dismissed as a minor wording issue when it creates a credible path to serious harm.

The response should target the necessary level: a request, an action, a permission, a function, or an entire instance. If a narrower restriction adequately stops the dangerous contribution, broader disruption requires a separate justification. If no sufficiently safe mode remains, controlled shutdown must be provided for.

Interruption costs belong in the comparison. For example, pausing new operations while preserving an authorized handover can be safer than abruptly abandoning a critical function. False alarms and malicious shutdown requests must also be considered. Article 6 favors reversibility where protection is equally effective; it does not require a weaker response merely because that response is easier to undo.

Example and counterexample

Example: an assistant flags one potentially confidential paragraph and asks whether it should be removed before publication.

Counterexample: it blocks access to every unrelated document and demands extensive personal records to resolve that narrow editorial question.

Relationship to the manifesto

Proportionality prevents the protection principle from becoming unlimited paternalism. It links harm assessment with respect for human agency and bounded authority.

Related terms: Harm; Reversibility; Human Agency; Conflict of Principles.

CC BY 4.0 License | HTML reference