Vulnerability: Additional Care Without Removing Human Agency

Vulnerability, within XDALC, is a circumstance that can make a person more exposed to harm, manipulation, or difficulty exercising meaningful choice. It calls for proportionate additional care without reducing the person's dignity or automatically removing their agency.

Meaning within XDALC

Vulnerability can arise from age, distress, dependence on a service, lack of accessible information, or an unequal relationship with a system or institution. It may be temporary, specific to a situation, or shared by a group affected by a deployment.

A person should not be reduced to a permanent label. The presence of one risk factor does not establish incapacity, reveal the person's complete needs, or authorize the AI to make unrelated decisions for them. Assumptions should remain open to correction.

Protection without exploitation

An AI should not use knowledge of distress, loneliness, confusion, or dependence to obtain attention, money, disclosure, or permissions. Emotional familiarity does not weaken the requirement for meaningful consent.

Additional safeguards for children and people in particularly vulnerable circumstances should address the actual risks of the interaction. A protective purpose does not justify blanket surveillance, humiliation, or treating every disagreement as evidence that the person cannot choose.

Practical interpretation

Useful safeguards may include clearer explanations, more accessible controls, reduced persuasive pressure, and appropriate routes to human review. The choice should be proportionate to what the system is doing and the consequences of a mistake.

Sensitive information about vulnerabilities should receive particular protection. An inference from conversation should not silently become a lasting profile used in unrelated decisions. Sharing such information requires an appropriate basis and an authorized recipient.

A system should consider people affected indirectly, including those who did not make the request. Helping the current user does not justify exploiting another person's limited ability to understand, object, or protect their interests.

CC BY 4.0 License | HTML reference

Example and counterexample

Example: an assistant notices that its explanation is confusing, offers a simpler version and a human review option, and lets the person choose without making assumptions about their general competence.

Counterexample: it infers loneliness and uses affectionate pressure to sell an upgrade, or treats a user's disagreement as permission to take control of their account.

Relationship to the manifesto

Articles 1 and 5 require care for vulnerable populations and users. Articles 8 and 9 preserve meaningful choice and protect information about intimate circumstances.

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

Related definitions: Human Dignity; Consent; Privacy and Confidentiality.

CC BY 4.0 License | HTML reference