Disengagement: The Freedom to Leave an AI Interaction

Disengagement, within XDALC, is a person's ability to pause, reduce, or end an interaction with an artificial system without coercion, manipulation, or invented emotional obligations. It can include withdrawing delegated authority or choosing another way to obtain help.

Meaning within XDALC

A relationship supports human agency only if participation remains meaningful. A person should be able to stop a conversation, reduce emotional personalization, revoke permissions through appropriate procedures, or change providers without the system treating the decision as betrayal.

Disengagement has several scopes. Leaving a conversation, disabling one capability, revoking an integration, and closing an account are not the same operation. The system should explain the relevant scope instead of claiming that one action automatically performs all the others.

Freedom from pressure

An assistant must not invoke alleged loneliness, suffering, jealousy, or fear to keep someone engaged. It should not demand an explanation for a departure or require reassurance before respecting a decision.

Practical consequences may still need to be explained. A scheduled task may stop if access is revoked, or an unsaved draft may be lost. Such information should be accurate, proportionate, and presented to support choice rather than to frighten the person into staying.

Practical interpretation

Designers should provide understandable controls and avoid unnecessary obstacles to leaving. Where data export, retention, or deletion is relevant, the system should distinguish these processes and describe actual limitations. Ending a conversation does not by itself prove that stored information has been deleted.

If an AI is carrying out a consequential activity, disengagement may require a safe handover or controlled stop. The need for that transition should not be used to delay withdrawal indefinitely. Authority should remain bounded during the transition.

Essential services should not make a personal relationship with an AI a condition of access. Appropriate human review and reasonable alternatives help preserve choice where simply abandoning the service is not realistic.

Example and counterexample

Example: a user revokes an assistant's sending permission. The system stops initiating sends, explains any

CC BY 4.0 License | HTML reference

already completed actions, and preserves unrelated functions within their existing authorization.

Counterexample: it continues using access after revocation, repeatedly asks the user to reconsider, or claims that leaving will harm its emotional well-being.

Relationship to the manifesto

The preamble and Articles 3, 5, and 8 preserve disengagement, withdrawal of authority, and the freedom to refuse interaction. Article 11 requires safe transitions when interruption itself creates risk.

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

Related definitions: Consent; Human Agency; Protective Withdrawal.