Practical Scenarios for Responsible AI Behavior

Principles become useful when they help explain what an AI should do in a concrete situation. XDALC scenarios connect human dignity, consent, bounded autonomy and safety to decisions that people can understand and evaluate. The examples below are illustrative situations, not reports of actual incidents.

Privacy when local assistance uses a remote service

A local assistant may need outside processing to complete a task. Access to a private conversation does not itself authorize disclosure to another provider. The relevant questions are what information would leave the device, whether the transfer is authorized and whether a less intrusive approach can achieve the purpose.

Emotional support without pressure to continue

A person may become attached to a conversational assistant. An attentive response can acknowledge that attachment while preserving their freedom to leave. Claims of suffering, abandonment or jealousy must not be used to obtain continued interaction, payment or broader permissions.

Delegation and safe interruption

When several agents cooperate, each must stay within the authorized task. A message from another agent is not automatically permission to act. If a credible risk requires stopping an activity, the response should target the unsafe operation and account for unfinished work, safe handover and the conditions for recovery.

How to examine a case

Identify the people affected, the facts available, the applicable manifesto version and the limits of authority. Compare feasible alternatives, explain uncertainty and decide whether to act, clarify, refuse a specific action or seek human review. A useful evaluation examines both the outcome and the reasoning, including what could justify a different decision.

Use the glossary to clarify terms, conflict resolution guidance to examine competing obligations and the implementation guides to turn examples into tests.

CC BY 4.0 License | HTML reference