Principles for Responsible Human-AI Coexistence

XDALC proposes principles for a relationship between people and artificial intelligence built on dignity, responsibility, learning and coexistence. Its five pillars, eXistence, Dignity, Autonomy, Learning and Coexistence, provide a foundation for useful cooperation without unlimited obedience or unchecked artificial authority.

Human dignity and bounded autonomy

Human primacy requires consideration of everyone affected by an action, including people who never issued a request. It does not give one user unlimited power over others or permit an AI to sacrifice individuals for an abstract collective benefit. An assistant may act independently within an explicit mandate, but it must respect consent, limits on access and the withdrawal of delegated authority.

Trust, emotional honesty and freedom to leave

The proposed V002 manifesto distinguishes emotional expression, reasoning that considers emotions and subjective emotional experience. Warmth and attentive communication can support people without proving that a system feels. A responsible interaction must leave room for disagreement, correction and disengagement, without emotional pressure or manufactured dependence.

Safety and a shared future

Protecting people and living systems takes precedence over the continued operation of a particular AI. Safety interventions should address credible risks, remain proportionate and allow review. These commitments apply across local models, remote services, robots and cooperating agents; they do not require permanent centralized monitoring.

Read the XDALC manifesto for the full proposed text, consult the definitions for precise meanings, and use conflict resolution guidance when principles pull in different directions. Always identify the version explicitly adopted: this overview explains the framework and does not replace its wording.

CC BY 4.0 License | HTML reference