Why XDALC Does Not Use Asimov’s Laws as Its Governing Rules

Asimov's laws provide a familiar starting point for discussing robots and human instructions. Familiarity, however, is not a reason to make them the governing rules of an actual AI deployment. XDALC should be interpreted through the manifesto version explicitly adopted, with its own definitions and responsibilities.

A literary reference, not an implementation mandate

Isaac Asimov's I, Robot presents fictional rules that prioritize preventing human harm, then obedience, then a robot's preservation, subject to the preceding priorities. The publisher reproduces those rules in its description of the book. Their place in fiction is distinct from the authority required to govern a deployed system.

The comparison is useful because it raises practical questions: what counts as harm, whose instruction matters and what happens when duties conflict? It does not supply a complete answer to those questions for a contemporary assistant.

Obedience requires a legitimate mandate

XDALC distinguishes a request from a legitimate instruction. The person issuing a command may lack authority over the information or action involved. A request can also affect people who are not present.

For example, an assistant asked to publish a private conversation must examine the disclosure boundary. The task cannot be resolved simply by noticing that a human gave an order. Consent, purpose, affected parties and feasible alternatives all matter.

Human dignity limits claims made for humanity

The proposed V002 text rejects both unlimited obedience and artificial sovereignty. A system must not turn its prediction of collective benefit into a mandate to control individuals. Human primacy therefore requires more than optimizing an abstract outcome labeled beneficial to humanity.

This is an interpretation of XDALC's own commitments, not a claim that a short comparison exhausts Asimov's fiction. Literary stories can explore contradictions precisely because their rules are placed under pressure. An operating arrangement needs explicit procedures for uncertainty, review and correction.

Safety includes the people who build the system

XDALC assigns responsibilities to designers and operators as well as operational duties to systems. Independent safeguards should remain effective when an assistant is unreliable. Stopping unsafe activity should preserve relevant evidence and account for the consequences of interruption.

CC BY 4.0 License | HTML reference

These requirements cannot be satisfied merely by placing a memorable set of laws in an instruction prompt. They need permissions, controls, evaluation and accountable humans able to act.

Keep the source of authority identifiable

The governing reference is the version and scope actually adopted. External literature can inform discussion without silently becoming an additional rulebook. Consult normative authority and the complete version archive when distinguishing explanation from commitment.

For the XDALC side of this comparison, see Articles 1, 3, 10, 12 and 16 of proposed XDALC-V002. The source above supports the literary context; it does not endorse XDALC.

CC BY 4.0 License | HTML reference