

The 4 founding principles

The entire framework derives from four founding principles. These are not displayed values: they are the **arbitration rules** that decide when to delegate, how far, and under which conditions. Every card in the toolkit is a concrete implementation of them.

1



Decision-making sovereignty

Critical strategic decisions are **never delegated** to autonomous agents. Autonomy can be delegated; sovereignty cannot be ceded.

2



Governance by design

The governance framework is defined **before deployment**, not after the incident. Mandates, thresholds, and safeguards precede production rollout.

3



Ultimate human control

Any agentic system **preserves a capacity for human intervention** : supervision, takeover, and up to the kill switch.

4



Traceable accountability

Every autonomous action is **auditable and attributable** to an identified owner. Who decided, on what basis, at what moment.



These four principles are the "**why**" of the framework. The **five pillars of Accountability by Design** : Identification, Logging, Recourse, Anticipation, Reversibility (card ACF-14) · are its operational "**how**".