

Safeguards for Bidirectional Transitions

Intent

Forward: add automation and transparency only when preconditions are met, in reversible steps.

Backward: remove automation only via time-boxed procedures with proofs and auto-reversion.

Preconditions & Triggers

Forward: legal/civic readiness, quality metrics, public consultation. Backward: emergency triggers, integrity threats, harm metric spikes.

Public Change Ledger

Append-only, hash-chained entries with dual signatures. Auto-reversion if signatures missing after 7 days.

Forward Path

Start with low-risk automation; contextualization by default; human-validated automation; transparent governance; ratchet with rollback.

Backward Path

Freeze sensitive automation; declare time-boxed manual mode; minimal necessary restrictions; oversight continuity; auto-reversion.

Anti-Abuse Guards

Quorum for deep changes, red team drills, risk sharding, version provenance.

User Protections

Appeals with SLA, downloadable case packets, data portability, point-of-impact explanations.

Governance & Runbook

Defined roles, checklists before/during/after shifts, and resilient metrics.