



OBJECTIVE

Ensure that every agentic system embeds, **from inception**, the five pillars of operational accountability — turning a doctrinal principle into an **auditable and enforceable organizational mechanism**.

1 DOCTRINAL LINEAGE

Responsibility-by-design (Frank Pasquale, *New Laws of Robotics*, Harvard, 2020) extends two established principles of responsible design:

CONCEPT	ORIGIN	LEGAL FRAMEWORK
Privacy by Design	Ann Cavoukian (Canada, 1995)	GDPR art. 25 (2018)
Security by Design	NIST · OWASP (2000s)	ISO 27001 · NIS2
Responsibility by Design	Frank Pasquale (Harvard, 2020)	AI Act art. 9 & 14 · DORA art. 5-8

2 THE FIVE OPERATIONAL PILLARS

IDENTIFICATION

Creators, controllers and owners publicly identifiable — an enforceable chain of accountability.

AUDIT LOGS

Append-only, timestamped and signed records — impossible for the operator to erase.

REDRESS

Users can always challenge decisions before a named human — timeframe and process defined in the code.

ANTICIPATION

Ex-ante mapping of drift scenarios — controls embedded in design, not bolted on post-incident.

REVERSIBILITY

Capability to disable, correct or withdraw the system — organizational and temporal cost defined upfront.



3 INTEGRATION WITH THE ACF TOOLKIT

Responsibility-by-design does not create an isolated mechanism — it **integrates** the toolkit's existing cards within a unified and enforceable framework:

PILLAR	ACF CARDS MOBILIZED	OPERATIONAL PRACTICE
Identification	ACF-03 · ACF-12	4 accountable signatures (business, technical, compliance, DDAO — Delegated Decision Agent Officer)
Audit Logs	ACF-08	Agentic Decision Register — append-only, timestamped, signed
Redress	ACF-05 · ACF-09	Review procedure + documented corrective action plan
Anticipation	ACF-02 · ACF-11	Criticality matrix + agentic risk assessment
Reversibility	ACF-06	Kill switch — 5-step shutdown procedure with defined SLA

4 PASQUALE'S FOUR LAWS — DOCTRINAL RECAP

NEW LAWS OF ROBOTICS — HARVARD, 2020

- 1 AI must **COMPLEMENT** professionals, not replace them.
- 2 AI must not **COUNTERFEIT HUMANITY** — no agent passing itself off as a human.
- 3 AI must not **INTENSIFY ZERO-SUM ARMS RACES**.
- 4 AI must always **DISCLOSE THE IDENTITY** of its creators, controllers and owners.

The **4th law** directly grounds the **Identification** pillar: without transparency on accountable parties, no chain of responsibility is enforceable. It is the foundation of all responsibility-by-design.



5 SELF-ASSESSMENT CHECKLIST

PAGE 3/4

To be completed by the DDAO with the team accountable for the system. **RbD-compliant = all pillars validated.**

Otherwise, trigger an ACF-09 action plan.

QUESTION	YES	PART.	NO
PILLAR 1 — IDENTIFICATION			
Does the system publicly disclose its creators, controllers and owners?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Are the accountable parties legally enforceable (4 ACF-03 Constitution signatures)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PILLAR 2 — AUDIT LOGS			
Are decisions traced in an append-only and signed register (ACF-08)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Can the system operator modify or erase these records? (<i>expected answer: NO</i>)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PILLAR 3 — REDRESS			
Can the user challenge a decision before a named human?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Are the timeframe and redress process defined from inception?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PILLAR 4 — ANTICIPATION			
Have drift scenarios been mapped ex-ante (ACF-02 + ACF-11)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Are countermeasures embedded in the design, not bolted on post-incident?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
PILLAR 5 — REVERSIBILITY			
Does a Kill switch (ACF-06) exist and is it tested periodically?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Are the organizational and temporal costs of rollback documented?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6 BEST PRACTICES & WARNING SIGNS

✓ BEST PRACTICES

- ✓ Have the ACF-03 Constitution signed by 4 named accountable parties
- ✓ Weekly audit of the ACF-08 Register for 90 days post-deployment
- ✓ Quarterly Kill switch testing (ACF-06) with report to the Board
- ✓ Documented redress pathway accessible within the product
- ✓ Update the ACF-11 risk assessment at every major evolution

⚠ WARNING SIGNS

- ! DDAO role combined with AI product owner (conflict of loyalty)
- ! Logs auditable only by the operator (breaks Pillar 2)
- ! Redress only buried in the T&Cs, not within the product
- ! Anticipation treated as a formality rather than as design
- ! Kill switch never tested — fictional practice, not enforceable



7 THE DDAO'S ROLE IN RBD

The **DDAO** organizationally embodies responsibility-by-design. They uphold the 5 pillars as non-negotiable requirements and hold a **reasoned right of objection** over any deployment that fails to meet them. Their reporting line preserves their independence.

RBD PILLAR	CONCRETE DDAO ACTION
1 · Identification	Verifies that the ACF-03 Constitution and the ACF-12 Mandate are signed before any deployment.
2 · Audit Logs	Audits the ACF-08 Register monthly — detects anomalies, suspends operations if needed.
3 · Redress	Safeguards the user challenge procedure and personally handles escalations.
4 · Anticipation	Validates the ACF-02 criticality matrix and the ACF-11 risk assessment ex-ante.
5 · Reversibility	Tests the ACF-06 Kill switch quarterly and reports to the Board or Audit Committee.

8 VALIDATION & SIGNATURES

ROLE	NAME	DATE	SIGNATURE
Business lead			
Technical lead			
Compliance Officer			
DDAO			

Related cards — ACF-02 Criticality matrix · ACF-03 Agentic Constitution · ACF-05 Supervision & governance · ACF-06 Kill switch · ACF-08 Agentic Decision Register · ACF-09 Action plan & improvement · ACF-11 Agentic risk assessment · ACF-12 Agent Mandate.

"Regulators are going to have to demand responsibility by design. That will entail mandating certain audit logs to be hardcoded, or licensing practices that explicitly contemplate problematic outcomes."

— Frank Pasquale, *New Laws of Robotics*, Harvard, 2020