



OBJECTIF

Garantir que tout système agentique intègre, **dès sa conception**, les cinq piliers de la responsabilité opérationnelle - pour passer d'un principe doctrinal à un **dispositif organisationnel auditable et opposable**.

1 FILIATION DOCTRINALE

La **responsibility-by-design** (Frank Pasquale, *New Laws of Robotics*, Harvard, 2020) prolonge deux principes établis du design responsable :

CONCEPT	ORIGINE	CADRE LÉGAL
Privacy by Design	Ann Cavoukian (Canada, 1995)	RGPD art. 25 (2018)
Security by Design	NIST · OWASP (années 2000)	ISO 27001 · NIS2
Responsibility by Design	Frank Pasquale (Harvard, 2020)	AI Act art. 9 & 14 · DORA art. 5-8

2 LES CINQ PILIERS OPÉRATIONNELS



IDENTIFICATION

Créateurs, contrôleurs et propriétaires publiquement identifiables - chaîne de responsabilité opposable.



AUDIT LOGS

Traces append-only, horodatées et signées - impossibles à effacer par l'opérateur.



RECOURS

L'utilisateur peut toujours contester auprès d'un humain identifié - délai et processus définis dès le code.



ANTICIPATION

Cartographie ex-ante des scénarios de dérive - contrôles intégrés au design, pas ajoutés post-incident.



RÉVERSIBILITÉ

Capacité de désactiver, corriger ou retirer le système - coût organisationnel et temporel défini.



3 ARTICULATION AVEC LE TOOLKIT ACF

La responsibility-by-design ne crée pas un dispositif isolé - elle **articule** les fiches existantes du toolkit dans un cadre unifié et opposable :

PILIER	FICHES ACF MOBILISÉES	PRATIQUE OPÉRATIONNELLE
Identification	ACF-04 · ACF-13	4 signatures responsables (métier, technique, compliance, DDAO - Delegated Decision Agent Officer)
Audit Logs	ACF-09	Registre des décisions agentiques - append-only, horodaté, signé
Recours	ACF-06 · ACF-10	Procédure de revue + plan d'action correcteur documenté
Anticipation	ACF-03 · ACF-12	Matrice de criticité + évaluation des risques agentiques
Réversibilité	ACF-07	Kill Switch - procédure d'arrêt en 5 étapes avec SLA défini

4 LES 4 LOIS DE PASQUALE - RAPPEL DOCTRINAL

NEW LAWS OF ROBOTICS - HARVARD, 2020

- 1 L'IA doit **COMPLÉTER** les professionnels, pas les remplacer.
- 2 L'IA ne doit pas **CONTREFAIRE L'HUMANITÉ** - pas d'agent se faisant passer pour un humain.
- 3 L'IA ne doit pas **INTENSIFIER LA COURSE AUX ARMEMENTS** à somme nulle.
- 4 L'IA doit toujours **INDIQUER L'IDENTITÉ** de ses créateurs, responsables et propriétaires.

La **4^e loi** fonde directement le pilier **Identification** : sans transparence sur les responsables, aucune chaîne de responsabilité n'est opposable. C'est le socle de toute la responsibility-by-design.



5



CHECKLIST D'AUTO-ÉVALUATION

À remplir par le DDAO avec l'équipe responsable du système. **Conforme RbD = tous les piliers validés.** À défaut, déclencher un plan d'action ACF-10.

QUESTION	OUI	PART.	NON
PILIER 1 - IDENTIFICATION			
Le système indique-t-il publiquement ses créateurs, contrôleurs et propriétaires ?	☐	☐	☐
Les responsables sont-ils opposables juridiquement (4 signatures Constitution ACF-04) ?	☐	☐	☐
PILIER 2 - AUDIT LOGS			
Les décisions sont-elles tracées dans un registre append-only et signé (ACF-09) ?	☐	☐	☐
L'opérateur du système peut-il modifier ou effacer ces traces ? (<i>réponse attendue : NON</i>)	☐	☐	☐
PILIER 3 - RECOURS			
L'utilisateur peut-il contester une décision auprès d'un humain identifié ?	☐	☐	☐
Le délai et le processus de recours sont-ils définis dès la conception ?	☐	☐	☐
PILIER 4 - ANTICIPATION			
Les scénarios de dérive ont-ils été cartographiés ex-ante (ACF-03 + ACF-12) ?	☐	☐	☐
Les contre-mesures sont-elles intégrées au design, pas ajoutées post-incident ?	☐	☐	☐
PILIER 5 - RÉVERSIBILITÉ			
Un Kill Switch (ACF-07) existe-t-il et est-il testé périodiquement ?	☐	☐	☐
Le coût organisationnel et temporel du retour arrière est-il documenté ?	☐	☐	☐



6



BONNES PRATIQUES & POINTS DE VIGILANCE

PAGE 4/5

✓ BONNES PRATIQUES

- ✓ Faire signer la Constitution ACF-04 par 4 responsables nommés
- ✓ Audit hebdomadaire du Registre ACF-09 pendant 90 j post-déploiement
- ✓ Test trimestriel du Kill Switch ACF-07 avec rapport au Conseil
- ✓ Voie de recours documentée et accessible dans le produit
- ✓ Mise à jour de l'évaluation des risques ACF-12 à chaque évolution majeure

⚠ POINTS DE VIGILANCE

- ! Cumul DDAO + responsable produit IA (conflit de loyauté)
- ! Logs auditable uniquement par l'opérateur (brise le pilier 2)
- ! Recours seulement dans les CGU, pas dans le produit
- ! Anticipation traitée comme formalité, pas comme design
- ! Kill Switch jamais testé - pratique fictive, non opposable

7



RÔLE DU DDAO DANS LA RBD

Le **DDAO** incarne organisationnellement la responsibility-by-design. Il porte les 5 piliers comme exigences non négociables et dispose d'un **droit d'opposition motivé** sur tout déploiement qui ne les respecte pas. Son rattachement hiérarchique préserve son indépendance.

PILIER RBD	ACTION CONCRÈTE DU DDAO
1 • Identification	Vérifie que la Constitution ACF-04 et le Mandat ACF-13 sont signés avant tout déploiement.
2 • Audit Logs	Audite mensuellement le Registre ACF-09 - détecte les anomalies, suspend si besoin.
3 • Recours	Garantit la procédure de contestation utilisateur, traite personnellement les escalades.
4 • Anticipation	Valide la matrice de criticité ACF-03 et l'évaluation des risques ACF-12 ex-ante.
5 • Réversibilité	Teste le Kill Switch ACF-07 trimestriellement, rapporte au Conseil ou Comité d'audit.



8 ✎ VALIDATION & SIGNATURES

RÔLE	NOM	DATE	SIGNATURE
Responsable métier			
Responsable technique			
Compliance Officer			
DDAO			

Fiches liées - ACF-03 Matrice de criticité · ACF-04 Constitution agentique · ACF-06 Supervision & gouvernance · ACF-07 Kill Switch · ACF-09 Registre des décisions agentiques · ACF-10 Plan d'action & amélioration · ACF-12 Évaluation des risques agentiques · ACF-13 Mandat d'agent.

« Les régulateurs vont devoir exiger une responsabilité dès la conception. Cela impliquera d'imposer certains audit logs codés en dur, ou des pratiques de licence qui prennent explicitement en compte les résultats problématiques. »

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