



OBJECTIF

Pouvoir **arrêter immédiatement** un agent ou un système agentique en cas de dérive, d'incident, de perte de contrôle ou de risque critique.

1 AGENT CONCERNÉ

NOM DE L'AGENT

ID / RÉFÉRENCE

MISSION PRINCIPALE

SYSTÈME / SERVICE ASSOCIÉ

PROPRIÉTAIRE (RESPONSABLE HUMAIN)

DATE DE CRÉATION

2 NIVEAU DE RISQUE DE L'AGENT

Évaluation du risque global en cas de dérive (cocher une option).

- | | |
|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Critique — impact majeur sur finances, clients, données ou conformité. | <input type="checkbox"/> Élevé — impact significatif sur opérations ou réputation. |
| <input type="checkbox"/> Modéré — impact limité et réversible. | <input type="checkbox"/> Faible — impact minimal. |

3 CONDITIONS D'ACTIVATION

Le Kill Switch doit être activé si l'une des conditions suivantes est remplie :

- | | |
|----------------------------------------------------------------|----------------------------------------------------------|
| <input type="checkbox"/> Dérive comportementale détectée | <input type="checkbox"/> Dépassement de limites définies |
| <input type="checkbox"/> Décision non conforme aux règles | <input type="checkbox"/> Impact financier anormal |
| <input type="checkbox"/> Plainte client / réclamation critique | <input type="checkbox"/> Violation de données ou d'accès |
| <input type="checkbox"/> Perte de contrôle / bug critique | <input type="checkbox"/> Instruction humaine explicite |
| <input type="checkbox"/> Autre (préciser) : | |



4

🕒

PROCÉDURE D'ARRÊT D'URGENCE

PAGE 2/4

ÉTAPE	ACTION	RESPONSABLE	DÉLAI / DÉTAIL
1. Détecter	Identifier le signal d'alerte ou l'incident		
2. Valider	Confirmer que la condition d'activation est remplie		
3. Activer	Déclencher le Kill Switch — action critique		
4. Isoler	Isoler l'agent et couper ses accès		
5. Contrôler	Contrôler l'arrêt et surveiller les effets		

5

⚡

ACTIONS AUTOMATIQUES À L'ACTIVATION

- ☐ Suspendre toutes les tâches en cours
- ☐ Bloquer les intégrations / API
- ☐ Notifier les responsables humains
- ☐ Révoquer les accès et tokens
- ☐ Mettre en lecture seule les données critiques
- ☐ Basculer en mode dégradé

6

🔔

NOTIFICATIONS

Qui prévenir, et par quel moyen ? (p. ex. responsable agent / DDAO, direction, DPO si données, métier)

RÔLE / PERSONNE	MOYEN (EMAIL / SMS / SLACK / AUTRE)



7 REVUE POST-INCIDENT

PAGE 3/4

À réaliser après chaque activation du Kill Switch.

- | | |
|----------------------------------------------------------------|-----------------------------------------------------------|
| <input type="checkbox"/> Analyse des causes racines | <input type="checkbox"/> Évaluation de l'impact réel |
| <input type="checkbox"/> Actions correctives à mettre en place | <input type="checkbox"/> Mise à jour des règles / limites |
| <input type="checkbox"/> Apprentissage et documentation | |

RESPONSABLE DE LA REVUE

DÉLAI DE REVUE

8 HISTORIQUE DES ACTIVATIONS

DATE / HEURE	AGENT	RAISON	ACTIVÉ PAR	DURÉE	IMPACT	REVUE

9 ACCÈS AU KILL SWITCH

Qui peut activer le Kill Switch ? (au moins 2 personnes recommandées)

NOM / RÔLE	CONTACT

- | | |
|-----------------------------------------------------|-----------------------------------------------------|
| <input type="checkbox"/> Méthode : interface dédiée | <input type="checkbox"/> Méthode : commande système |
|-----------------------------------------------------|-----------------------------------------------------|



10 TEST RÉGULIER

PAGE 4/4

Un Kill Switch jamais testé est un Kill Switch incertain. Planifier un test périodique en conditions réelles.

FRÉQUENCE (MENSUELLE / TRIMESTRIELLE / SEMESTRIELLE / ANNUELLE)

DERNIER TEST RÉALISÉ LE / PAR

RÉSULTAT DU TEST

PROCHAIN TEST PRÉVU LE

Un agent puissant sans Kill Switch est un risque, pas un atout.

Prévoir l'arrêt, c'est protéger la valeur et la confiance. — Voir **ACF-05 Supervision** et **ACF-11 Évaluation des Risques Agentiques**.



VALIDATION DU PROTOCOLE

RESPONSABLE AGENT (DDAO) — NOM / DATE / SIGNATURE

DIRECTION GÉNÉRALE — NOM / DATE / SIGNATURE