



Comité Français des Tests Logiciels

Association loi 1901
26-28 Rue de Londres
75009 Paris

<https://www.cftl.fr/>
@cftl



VERS UNE GOUVERNANCE ANTIFRAGILE

Décider à l'ère des collectifs humain - agent IA

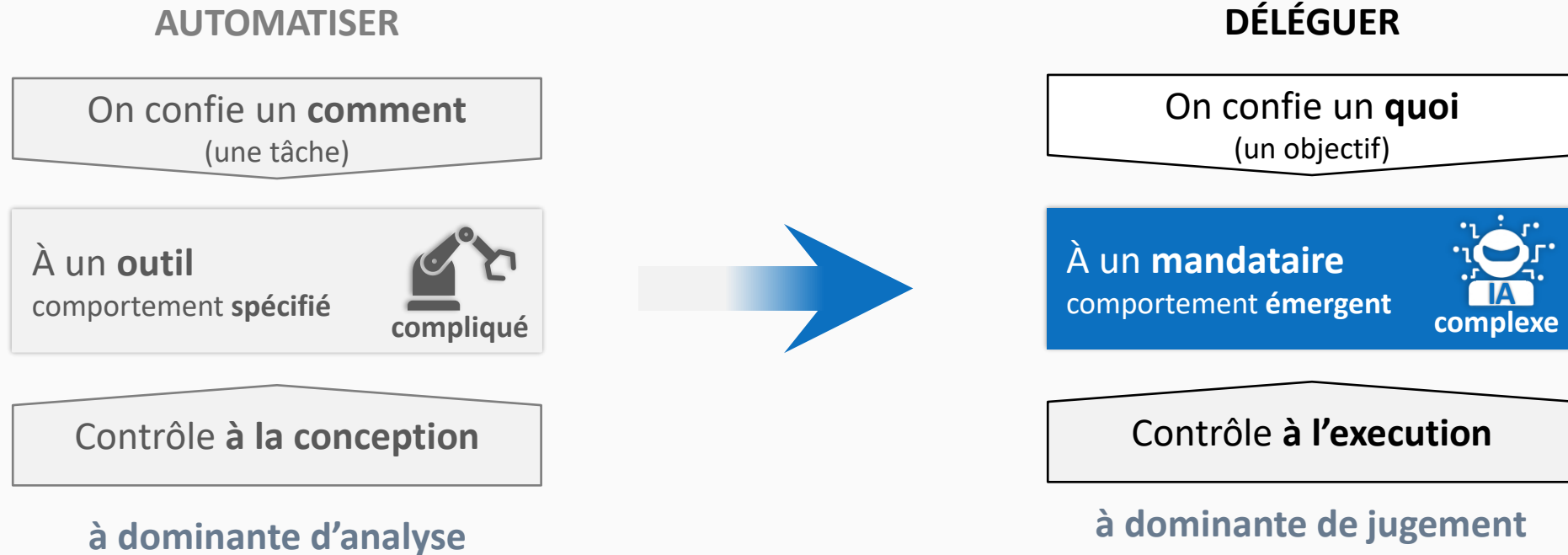
9 Juin 2026 - Paris, France

Présentation

Gabriel ROUSSEAU -- IBM

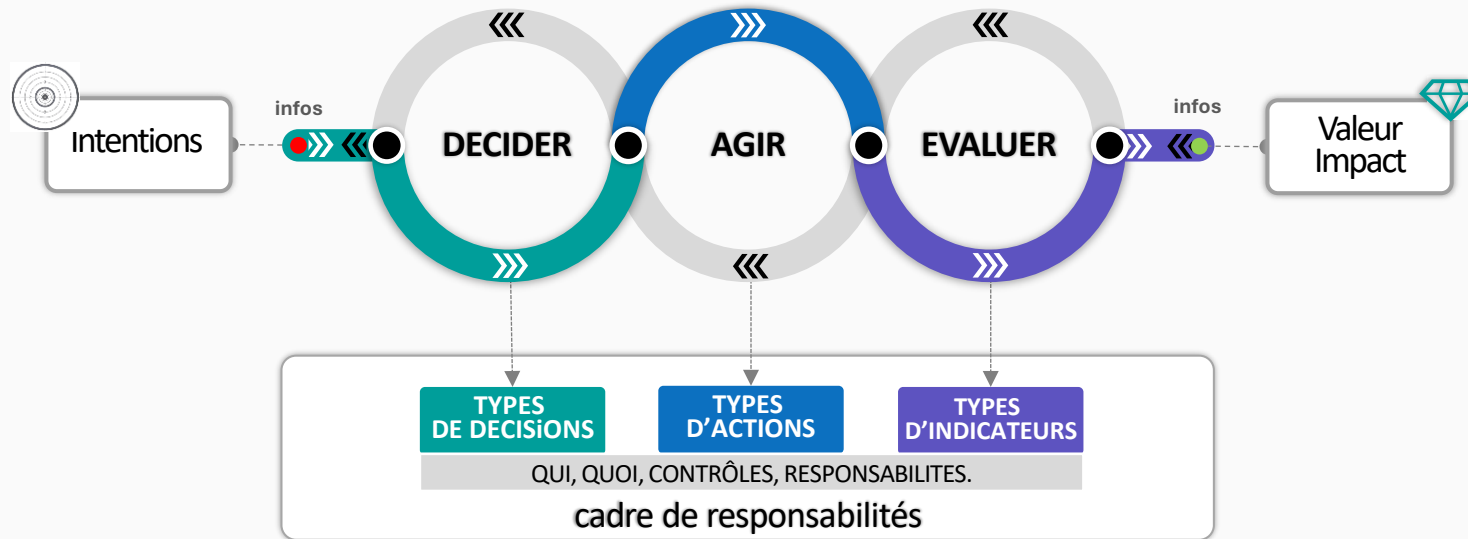
Karim HEDEOUD-PERROT -- IMPACTEEV

La rupture



On n'automatise plus, on délègue.
Et déléguer sans contractualiser, c'est juste espérer.

Le Cycle de Gouvernance



COMPLIQUÉ ← CONTEXTE DE GOUVERNANCE → COMPLEXE

surtout des **règles** ▶ logique de prescription
écart à la spécification ▶ **bug à corriger**

Aussi des **Boucles d'apprentissage** ▶ logique adaptation
écart à l'intention ▶ **dérive à comprendre**

Le complexe se gouverne par l'adaptation permanente et la capacité à juger.

Le Harnais

L'infrastructure qui rend la délégation gouvernable

*Ce que l'humain construit autour de l'agent pour qu'il puisse **agir pleinement dans le respect d'un cadre.***

01

Permissions

Ce que l'agent a le droit ou pas de faire.

02

Observabilité

Ce qu'on peut voir de l'activité de l'agent.

03

Points de contrôle

Là où l'humain peut intervenir.

04

Traces

Ce que l'agent laisse pour qu'on apprenne.

ces 4 dimensions sont les clauses d'un **contrat de délégation**
(objet, limites, conditions de révocation, critères d'évaluation).

Sans harnais, la gouvernance reste une intention. Avec harnais, **elle devient un dispositif.**

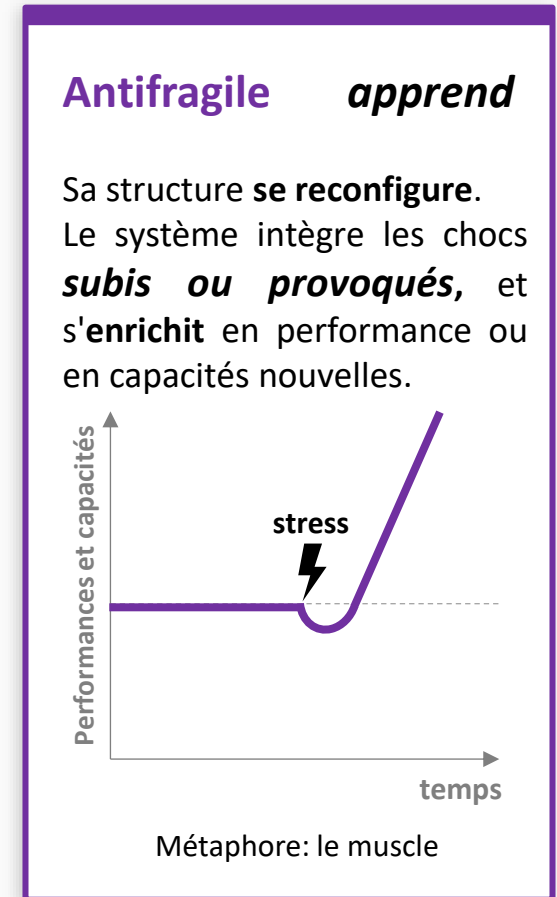
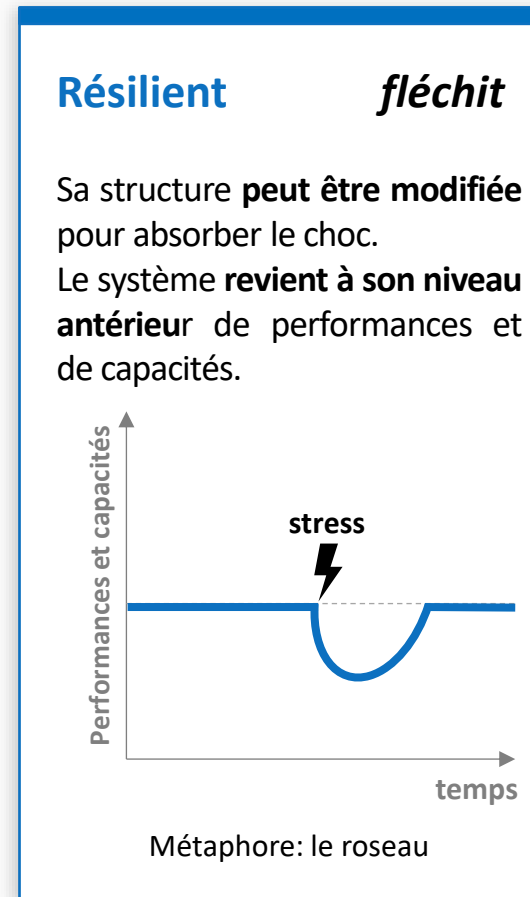
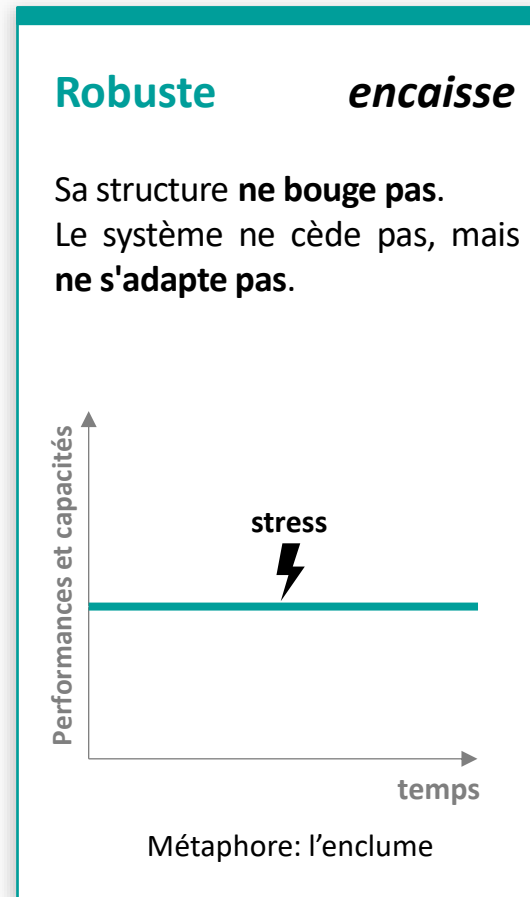
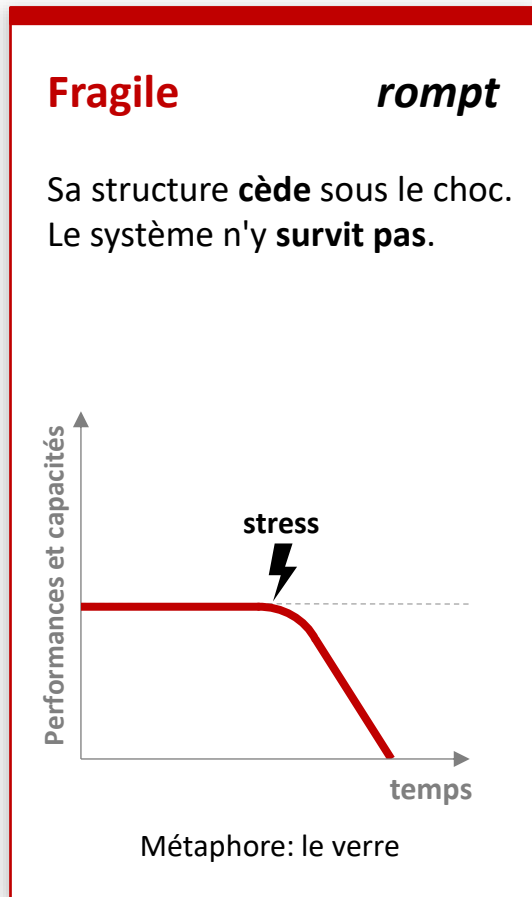
Ce que vous allez devoir tester a changé de nature.



Votre métier ne disparaît pas. Il monte d'un cran.

Du fragile à l'antifragile

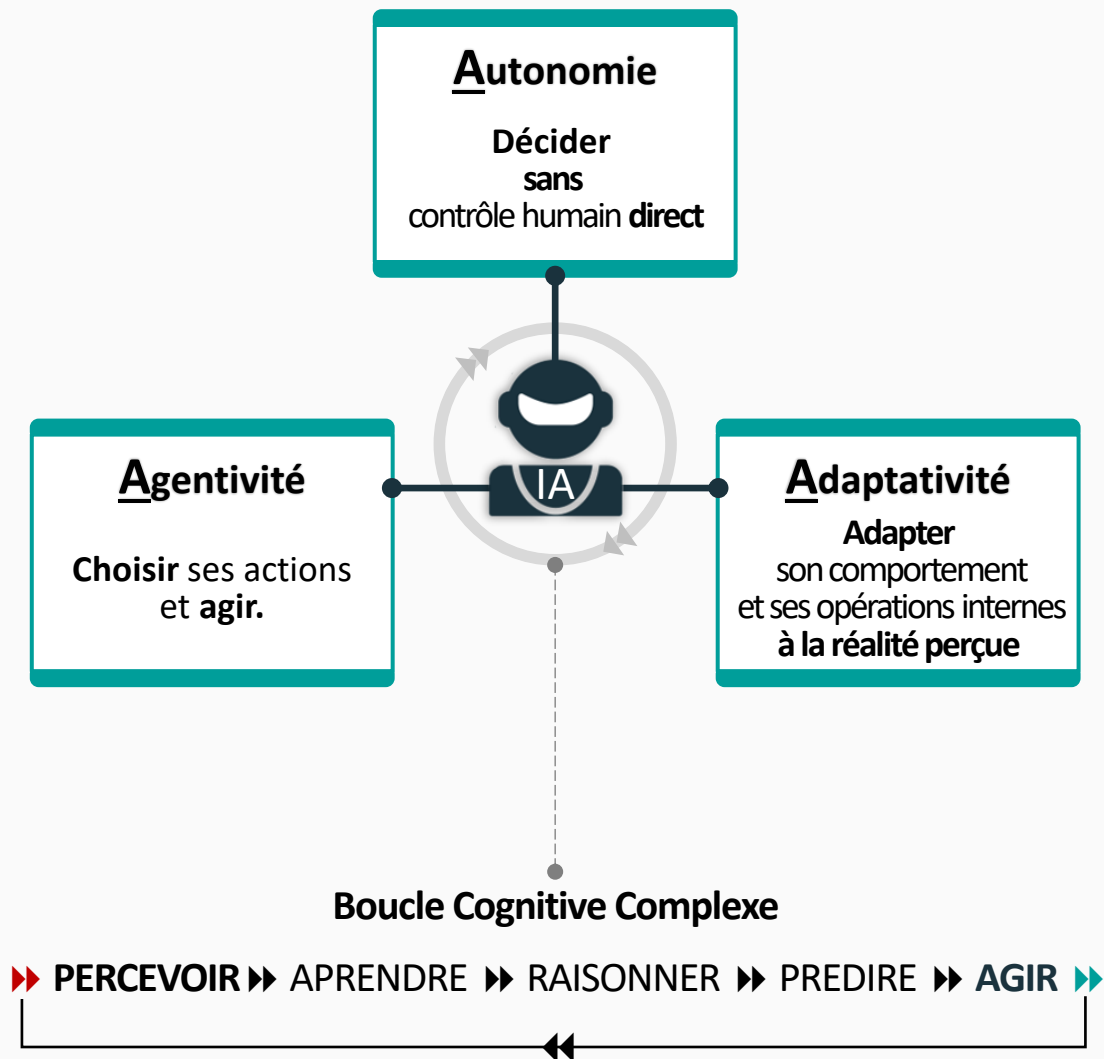
Réponse possible d'un système aux stressseurs (chocs, perturbations) en contexte d'incertitude



Provoquer les chocs pour apprendre, plutôt que d'attendre qu'ils s'imposent

Autonomie & dérive d'agents AI^{AAA}

Pourquoi la dérive est structurelle



Autonomie

Pour agir, un agent se construit un **modèle du monde** en cohérence avec la part de réel qu'il perçoit.

Incomplétude de la modélisation

Le réel est intrinsèquement **irréductible** à un modèle.

Dérive émergente

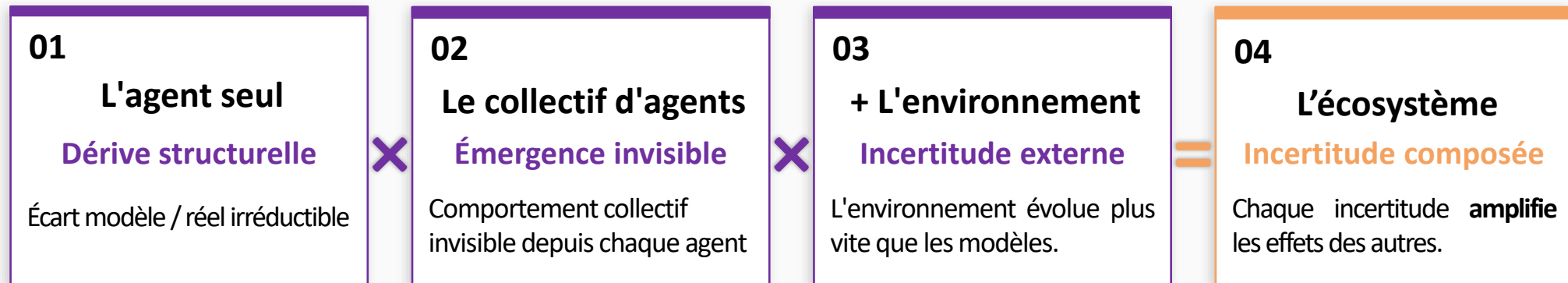
C'est l'**écart structurel** entre la cohérence du modèle de l'agent et la nature **irréductible** du réel.



la dérive n'est pas un bug,
c'est une conséquence
de ce qui rend l'agent capable et autonome.

Pourquoi une gouvernance antifragile pour un système agentique IA?

Plus qu'une montée en complexité, un changement de nature de l'incertitude.
"Pus d'agents, plus de dérives, mais c'est gérable avec plus de tests et plus de monitoring". Faux!!!!



L'incertitude est **constitutive** du système agentique et de son écosystème.
Chaque dérive *observée ou provoquée* est une opportunité d'apprentissage et de gain durable.

Mise en garde

Antifragilité ne veut pas dire improvisation.

Elle exige plus de rigueur qu'une gouvernance classique.

Provoquer des chocs contrôlés

Du désordre maîtrisé, pas subi

Revue post-incident structurées

Comprendre, pas seulement corriger

Des boucles de rétroaction formalisées

Apprendre par méthode, pas par hasard

C'est un changement d'état d'esprit, pas de vocabulaire.

3 piliers d'une gouvernance antifragile

01



Rendre les écarts visibles

La dérive silencieuse est l'ennemi de la gouvernance.

L'agent **doit signaler** son doute à chaque étape de son activité. Le harnais **dicte** ses droits: voir, dire, demander.

02



Organiser l'arbitrage humain

La zone d'escalade doit être balisée, pas improvisée.

Qui tranche, sur quoi, en combien de temps : à définir a priori, dans le **harnais**. L'**arbitrage** diffère selon que l'agent décide, agit ou évalue.

03



Faire de l'erreur une ressource

Corriger sans apprendre, c'est rester fragile.

Chaque dérive (**observée ou provoquée**) révèle une **hypothèse fautive**. Sa trace appelle une révision du contrat de délégation.

Trois questions pour votre organisation

1

Mes agents signalent-ils leur doute à chaque étape de leur activité?

Quand ils décident, agissent et évaluent, restent-ils silencieux quand ils doutent?

2

Ai-je défini qui arbitre quoi, et dans quel délai?

Pour décider, agir et évaluer, qui tranche, sur quels critères, en combien de temps?

3

Est-ce que j'apprends de chaque dérive, ou je la corrige pour qu'elle disparaisse?

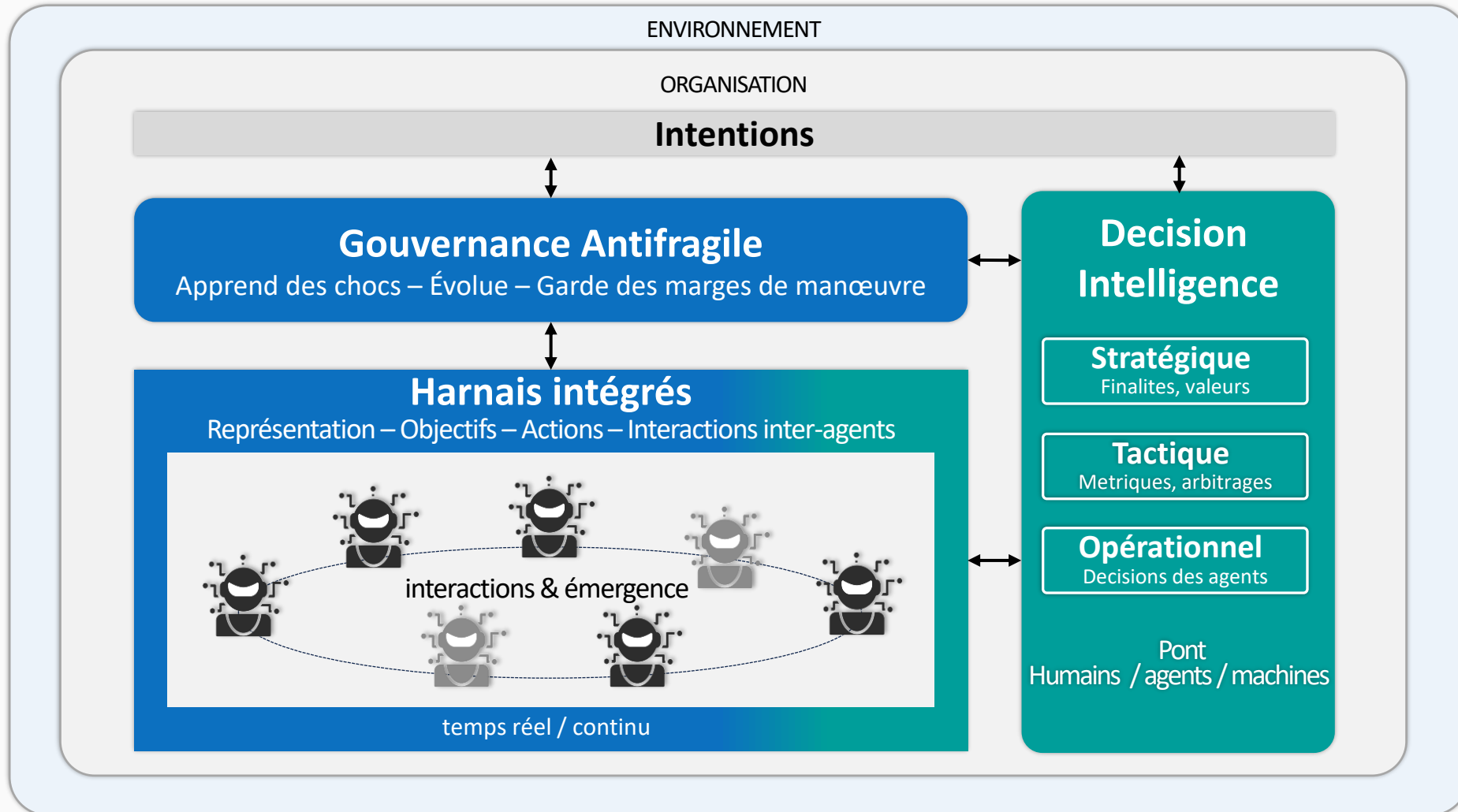
La dérive implique-t-elle de réviser le contrat de délégation, ou passe-t-elle sous silence?

Chacune question teste un pilier.

Une organisation qui peut y répondre **oui** a commencé à devenir **antifragile**.

Gouvernance et Decision Intelligence

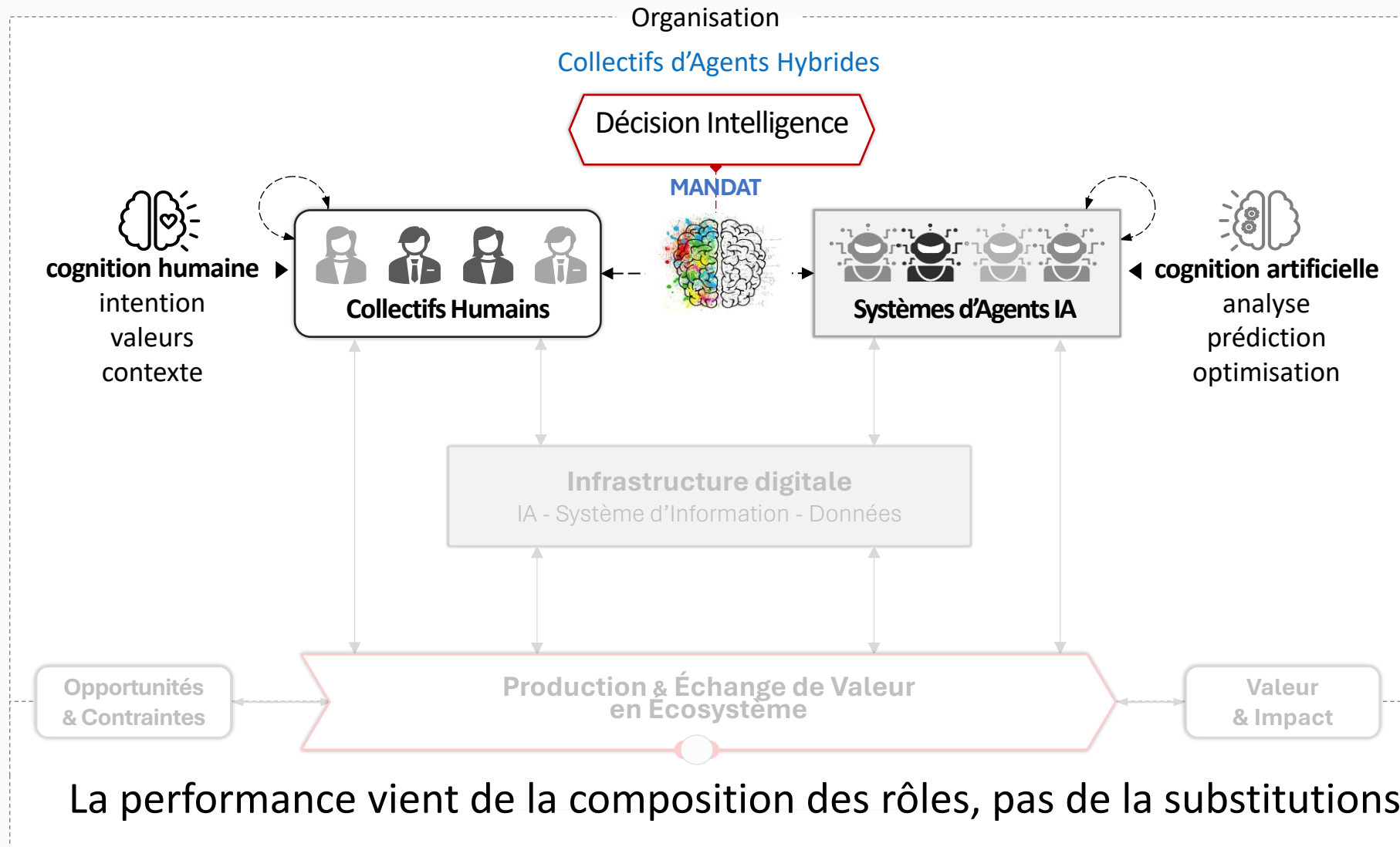
Deux disciplines complémentaires



La gouvernance donne le sens · la decision intelligence en donne les moyens · le collectif hybride compose l'ensemble.

Decision Intelligence

et Collectifs d'Agents Hybrides (Humains & IA)



Définir le MANDAT des décisions

6 dimensions pour déléguer des décisions

Mission

Définir l'objet de la délégation

Autonomie

Choisir le niveau de contrôle suffisant

Normes

Cartographier les règles et les politiques

Données

Auditabilité et intégration dans le **Harnais**

Arbitrage

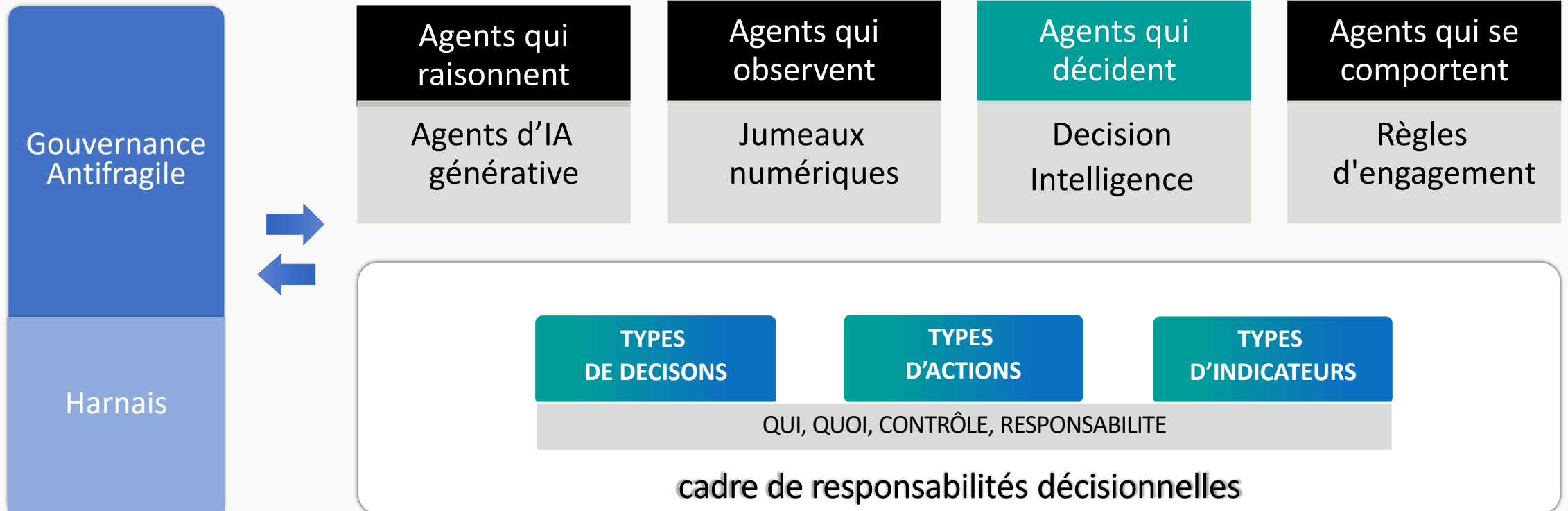
Baliser la zone d'escalade

Testabilité

Boucle d'amélioration décisionnelle **Antifragile**

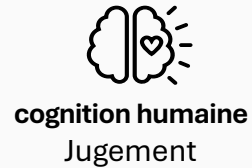
Mission: Définir l'objet de la délégation

Quelle décision confiez-vous vraiment ?



Autonomie: Choisir le niveau de contrôle suffisant

Adapter l'autonomie de la décision au contexte



Chaotique

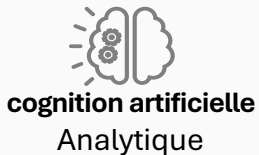
→ **Reprise humaine**

Jugement humain

Complexe

→ **Co-décider**

Jugement humain en dernier recours



Clair

→ **Automatiser ou pas !**

Analytique

Compliqué

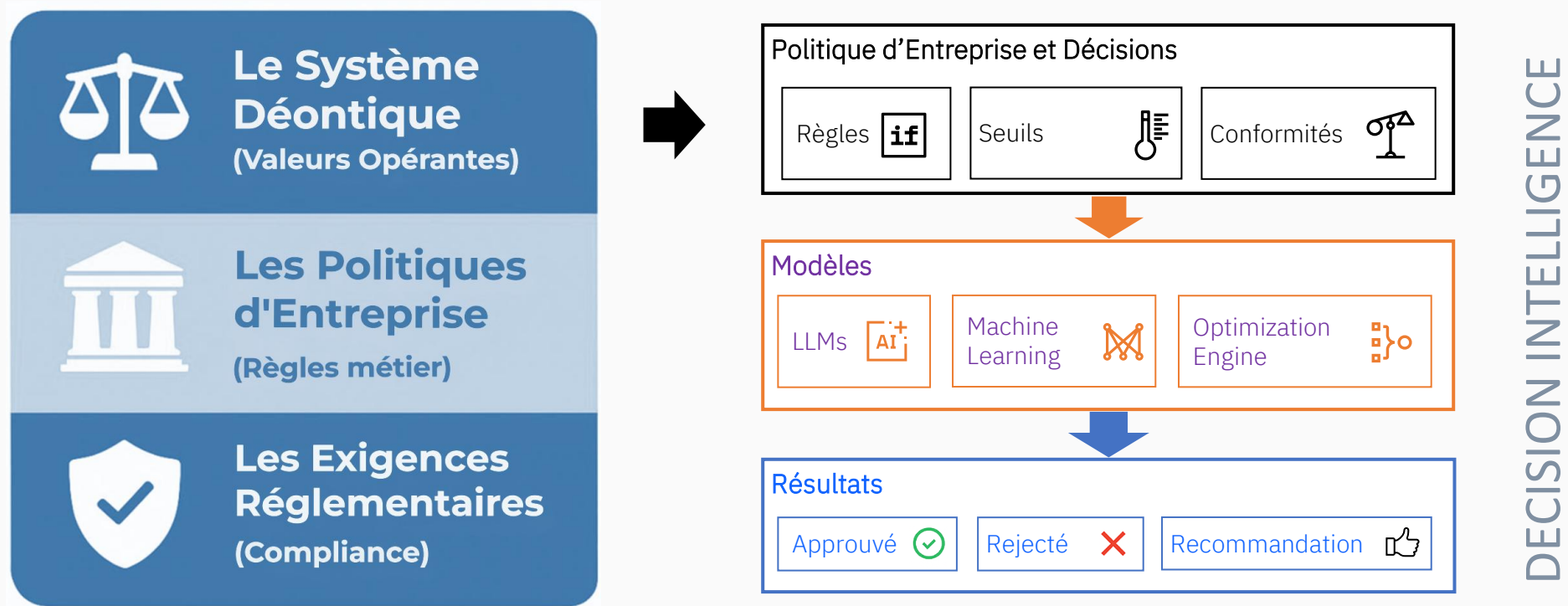
→ **Augmenter**

Analytique

Risque / irréversibilité ►

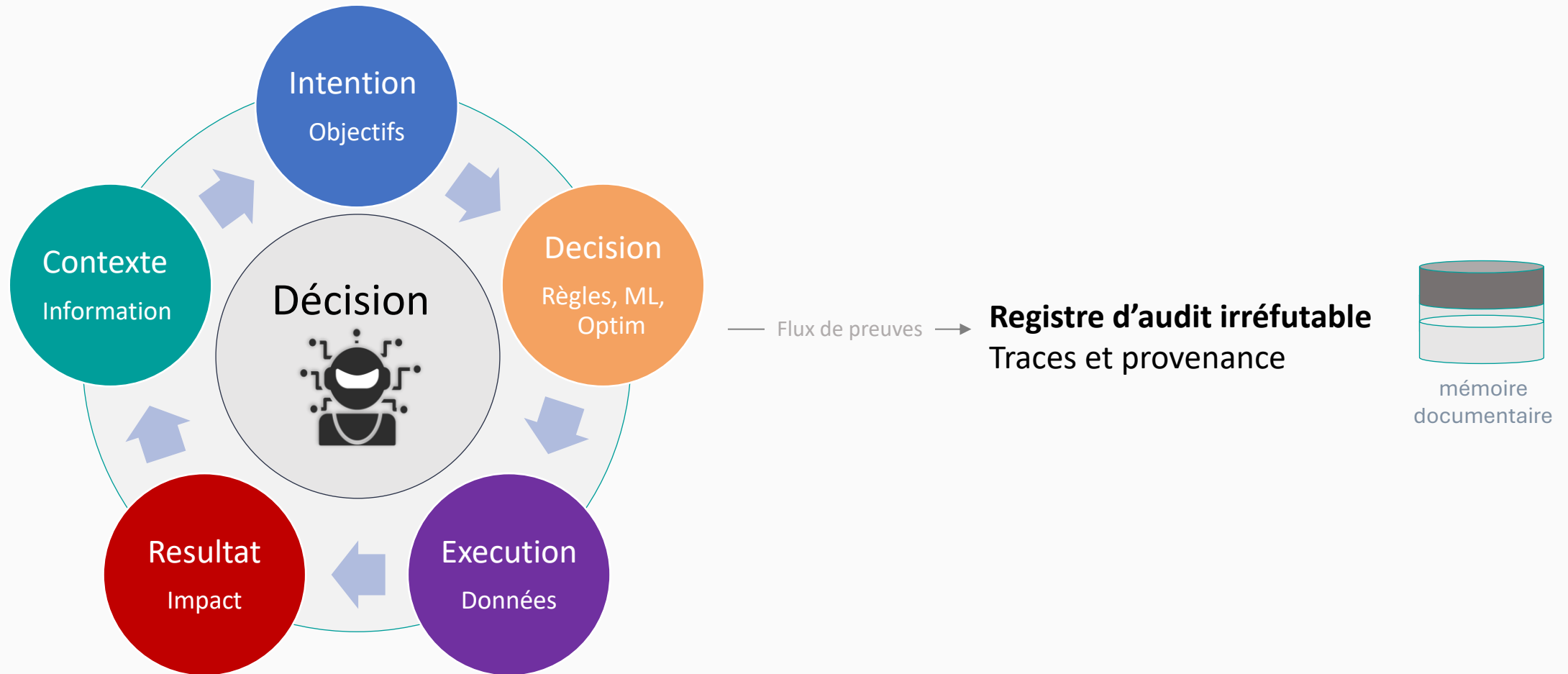
L'autonomie est un choix de risque

Normes organisationnelles: cartographier les règles et les politiques



*les politiques doivent être encodées par des **règles** pour garantir l'**auditabilité** des décisions.*

Données: Auditabilité et intégration dans le Harnais



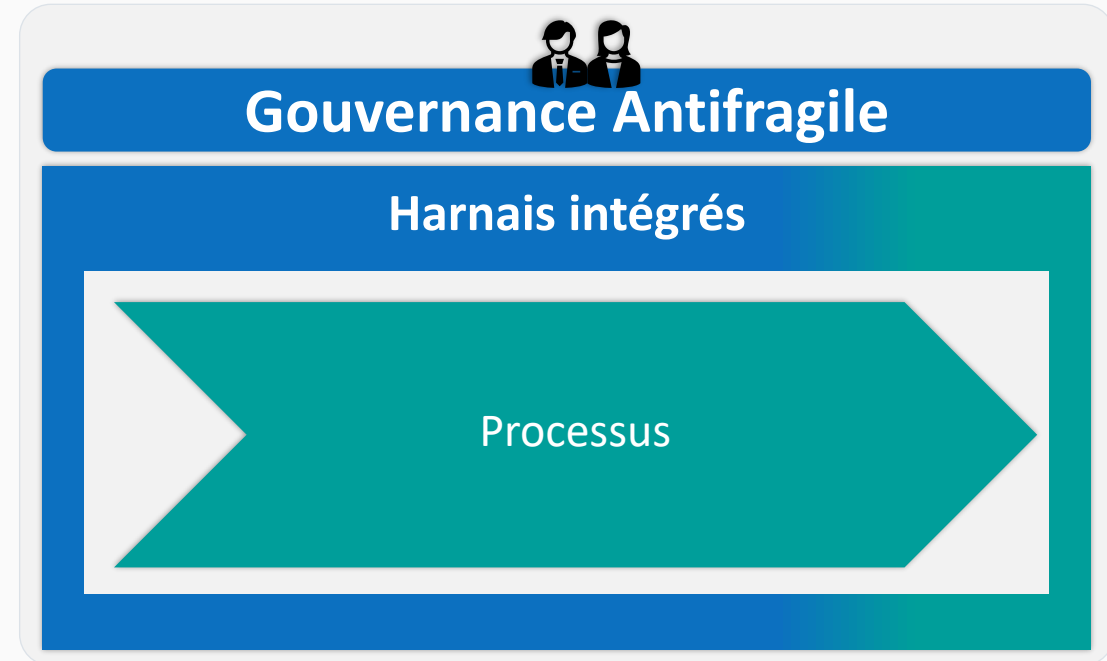
Infrastructure agentique pour agir en respectant le cadre

Arbitrage: Baliser la zone d'escalade

Replacer l'humain au dessus de la boucle



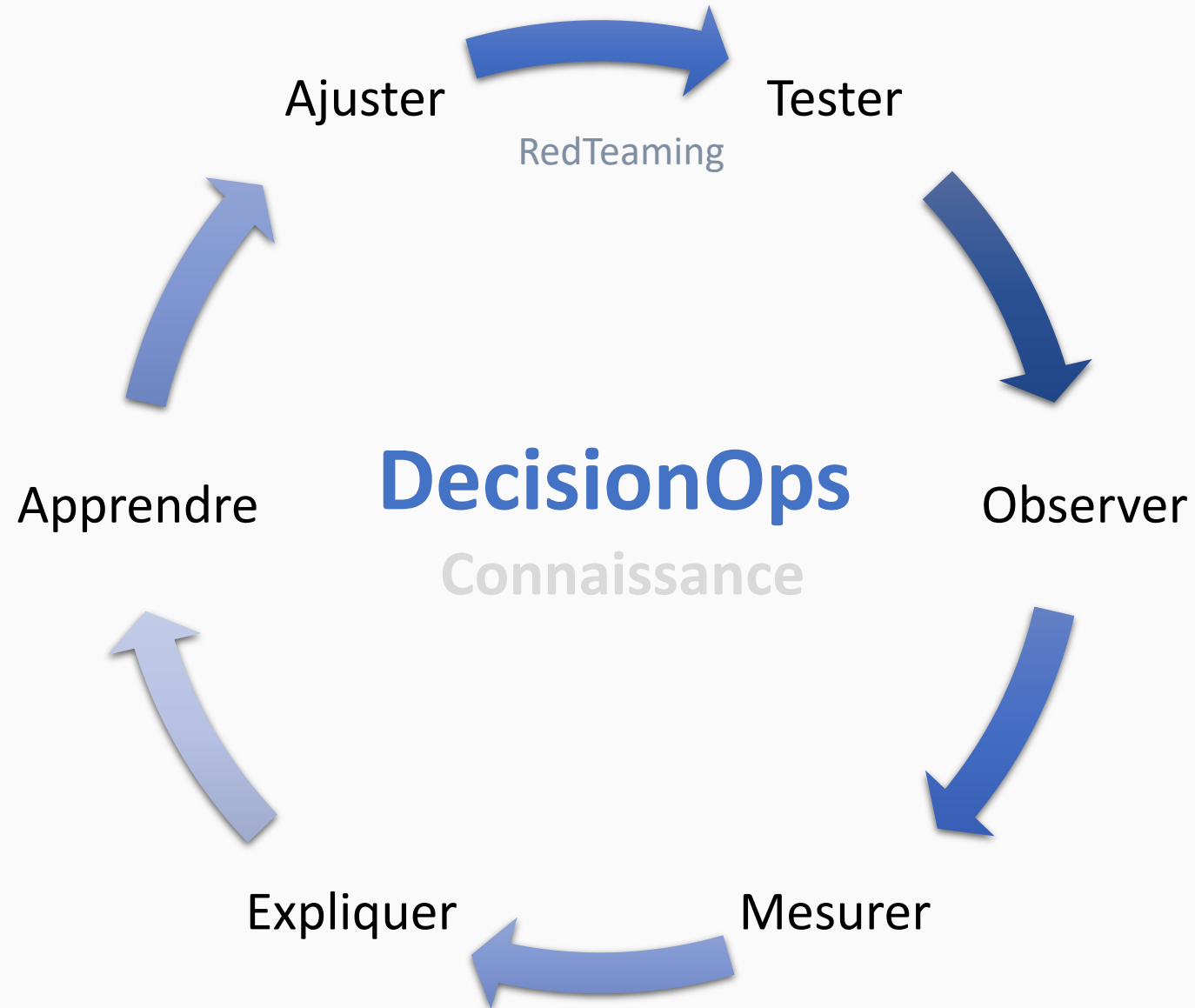
Humains paralysés



Humains au dessus de la boucle

L'humain est responsable: décide, agit et évalue

Testabilité: boucle d'amélioration décisionnelle Antifragile



Sans traces, pas d'apprentissage, pas d'Antifragilité

Où se déplace la valeur humaine?



L'impact n'a de sens qu'à l'aune de la volition qui l'a précédé.

Un résultat sans intention formulée n'est pas un impact, c'est un effet.

Votre métier ne disparaît pas. Il devient stratégique.

Principe de Jevons appliqué à l'IA

Rendre une ressource moins chère n'en réduit pas la consommation, ça l'augmente.

L'exécution devient quasi gratuite



La demande de supervision qualifiée explose

Plus vous déployer d'agents pour décider, agir, évaluer



Plus vous aurez d'écarts à juger.

Chaque dérive enrichit le système



Plus il grandit, plus il a besoin de vous.

Le testeur qui lit une dérive comme un signal sur la qualité d'une intention, celui-là est irremplaçable.

II.A. Modèle de Mandat Hybride

1. IDENTIFICATION EQUIPE	Nom de l'équipe hybride : Mission de l'équipe : Date de création du mandat : Version : Responsable de l'équipe (humain) :
2. COMPOSITION EQUIPE	Membres humains : - [Nom, rôle, responsabilités] - [Nom, rôle, responsabilités] Agents IA membres : - [Nom agent, type, rôle dans l'équipe] - [Nom agent, type, rôle dans l'équipe] Ratio : [X humains / Y agents]
3. MATRICE DE DECISION	TACHES RESERVEES AUX HUMAINS : - [Tâche 1 : justification] - [Tâche 2 : justification] TACHES DELEGUEES AUX AGENTS : -

Êtes-vous prêts à gouverner
ce que vous commencez
déjà à déléguer?

	objectifs - surcharge humains (agents insouffisants) - sous-utilisation agents (délégation insuffisante) Contact : [Responsable équipe] Délai : [temps] Action : [pause agents, réallocation tâches, etc.]
12. JOURNALISATION	Logs obligatoires : - Toutes tâches déléguées par humains à agents - Toutes explications agents vers humains - Tous désaccords H-A - Toutes décisions collectives H+A - Toutes modifications répartition tâches - Révision : [période] Accès : [qui peut consulter]
13. VALIDATION	Validé le : Par : [Responsable équipe humain] Prochaine révision : [date] Note : Révision à chaque changement composition équipe (arrivée/départ H ou A)



Comité Français des Tests Logiciels

Association loi 1901
26-28 Rue de Londres
75009 Paris

<https://www.cftl.fr/>
@cftl



***"On mesure l'intelligence d'un individu
à la quantité d'incertitudes
qu'il est capable de supporter."***

attribuée à E. KANT

Présentation
Gabriel ROUSSEAU -- IBM Karim HEDEOUD-PERROT -- IMPACTEEV