Intelligence artificielle

Vers une approche globale de gestion des risques

20/07/2025

TLP:CLEAR

PAP:CLEAR

Le projet en synthèse

La problématique : on ne sait plus par quel bout prendre le problème

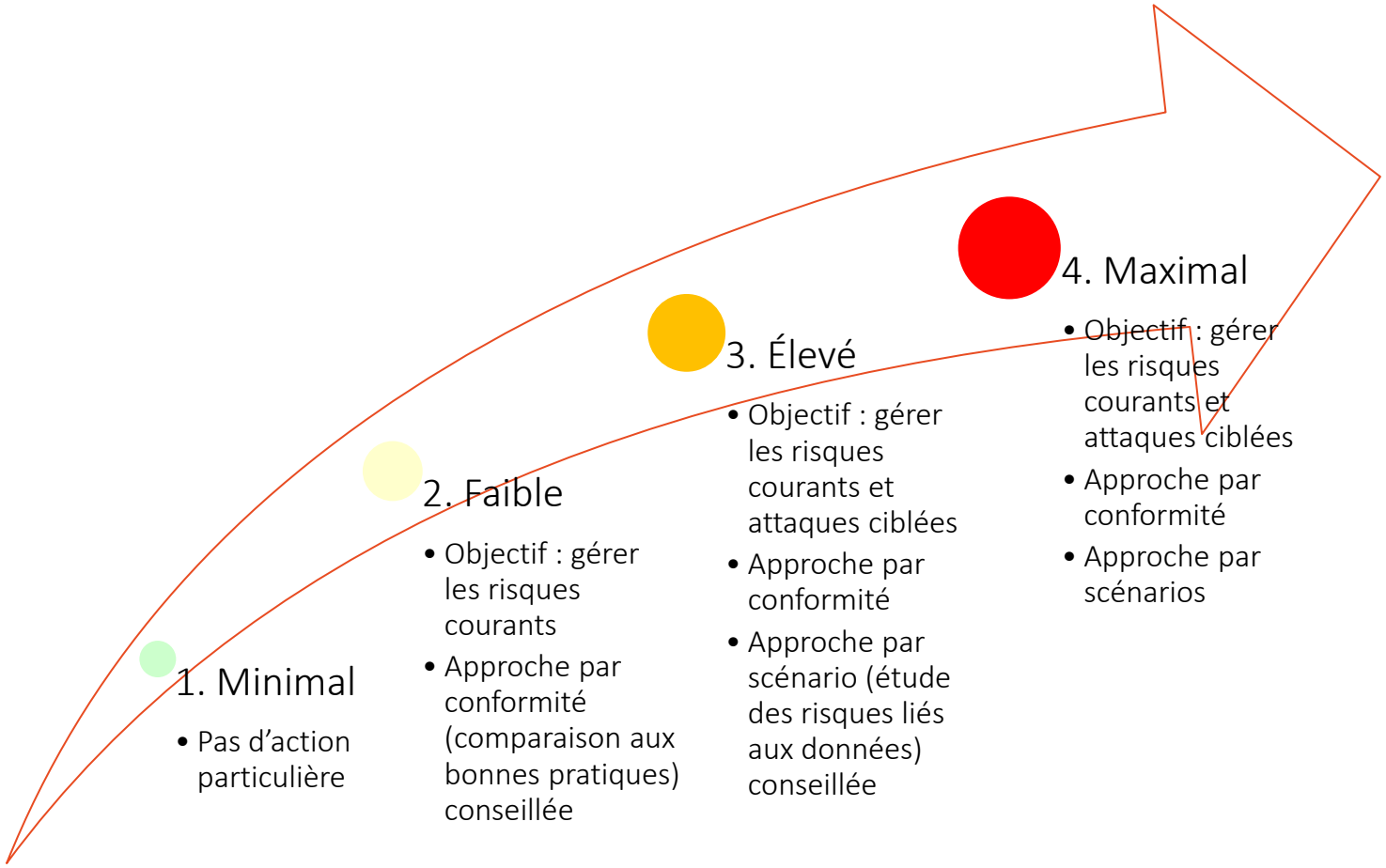
- **Une multitude de référentiels** et travaux en cours, chacun incomplet, tous incohérents !
- **Les institutions ont des domaines de compétences limités**
- **Les organisations ne segmentent pas** : elles doivent tout appliquer !
- **Un Règlement IA général en théorie, des projets de normes inapplicables en pratique**
  - Lobby rejetant avec force tous les travaux d'harmonisation réalisés depuis des dizaines d'années et ayant fait l'objet de consensus internationaux de tous les domaines au profit d'une approche américaine de sureté de dispositifs médicaux

L'objectif : expliquer par quel bout prendre le problème !

- **Une méthode de gestion des risques** de produits ou services qui reposent sur l'IA
- **Simple** !
- **Véritablement globale** (sécurité de l'information, protection de la vie privée, protection de l'environnement, etc.)
- **Basée sur les risques**
- **Harmonisée** et **cohérente**
- **Exploitant les référentiels existants** (ne réinventant pas !)
- **Flexible** et **intégrable** aux outils et procédures existants

La logique générale

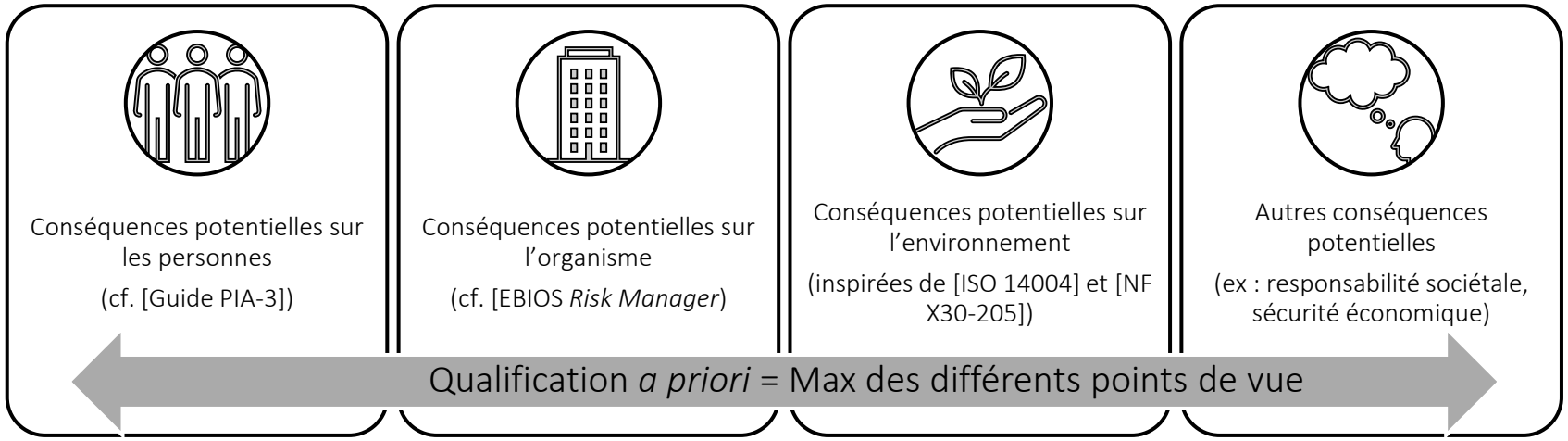
Une approche simple, globale, flexible, et proportionnée aux risques



Étape 1 : estimer le niveau de risque

Une échelle extensible pour proportionner l'étude aux risques

Niveau de risque
1. Minimal
2. Faible
3. Élevé
4. Maximal



Étape 2 : approche par conformité

L’objectif : gérer les risques courants

- Problème : aucune harmonisation
  - Les référentiels se ressemblent…
  - Ils emploient tous un classement légèrement différent
  - Les bonnes pratiques sont redondantes et incohérentes
- Solution : un référentiel pivot
  - Des critères de confiance harmonisés
  - Une synthèse des bonnes pratiques
  - Un lien vers les référentiels existants
- Avantages
  - Possibilité de filtrer
  - Utile (voire suffisant) pour homologuer
  - Valorisation des référentiels existants


Gouvernance responsable


Fiabilité et sûreté


Équité


Transparence


Sécurité des informations


Protection des droits et libertés


Maintenance et évolutivité


Interopérabilité


Respect de l’environnement

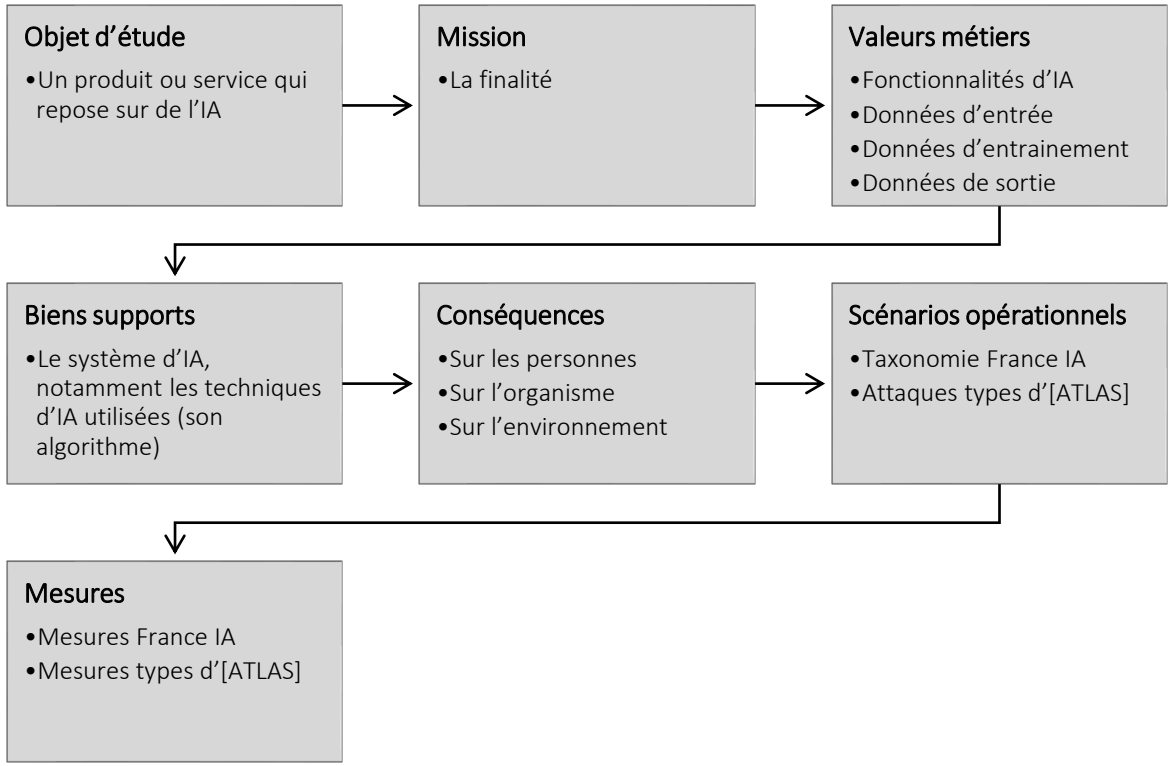

Accessibilité

Étape 3 : approche par scénarios

L'objectif : gérer les attaques ciblées

- Problème : pas d'approche définie
  - Pas de méthode harmonisée pour gérer les risques spécifiques à l'IA
- Solution : spécialiser sa méthode
  - Nos méthodes de gestion des risques de sécurité de l'information fonctionnent !
  - Prendre en compte les spécificités de l'IA : données, algorithmes, conséquences au-delà de l'organisme, attaques et mesures spécifiques
- Avantages
  - On n'ajoute pas de nouvel outil
  - On considère réellement les spécificités
  - Directement utile pour homologuer

Exemple : spécialisation d'EBIOS *Risk Manager*



Pour finir…

- Cette approche peut être utilisée comme **approche globale pour gérer les risques liés à l'IA**
- Elle peut également être utilisée comme **« *système de gestion des risques* » pour gérer les risques des « *systèmes d'IA à haut risque* »** (sans présomption de conformité)
- Elle sert aussi à **améliorer divers documents de référence et travaux en cours** ou à venir : [ISO/IEC 42001], [ISO/IEC 23894], [ISO/IEC 42005], [Guide de France IA], et même sans doute [ISO/IEC 27001] et [ISO/IEC 27005]
- **Vous pouvez contribuer** à l'améliorer et l'enrichir : https://github.com/matthieu-grall/ai

Ressources utilisées

- Images
  - Page de garde : Grid, par Magic Creative, de PIXABAY
  - Logo : Matthieu GRALL expert-conseil
  - Autres : photothèque de MICROSOFT Office

Informations de versions

Version	Action	Éditeur
28/06/2025 (v0.1)	Création de la présentation	Matthieu GRALL
20/07/2025 (v1.0)	Finalisation d'une première version complète de la présentation	Matthieu GRALL

Zoom sur le périmètre du [Règlement IA]

Des méprises possibles

- En apparence, le périmètre de gestion des risques du [Règlement IA] peut sembler limité à "*la santé, la sécurité et les droits fondamentaux*"
- Or, les seules formulations "*santé*", "*sécurité*" et "*droits fondamentaux*" mélangent des domaines de compétences, des causes et des conséquences
- Qui plus est, ces formulations sont :
  - ambigües, ex : traduction maladroite en "*sécurité*" au lieu de "sûreté" ;
  - redondantes, ex : les conséquences considérées liées à santé et sécurité se recouvrent ;
  - incomplètes : des défauts d'accessibilité peuvent engendrer des conséquences sur les "*droits fondamentaux*", des risques de sécurité de l'information peuvent engendrer des conséquences sur "*la santé, la sécurité ou les droits fondamentaux*", etc.

Les périmètres considérés

Un projet qui se veut très large

- Le périmètre du [Règlement IA] est en fait déjà large
  - **Tous les produits et les services qui reposent sur des systèmes d'IA** (certains imaginent qu'il ne traite que de produits du fait que certains articles ne concernent que les produits !)
  - **Toutes les causes de risques** : problèmes de sûreté, incidents de sécurité de l'information d'origine accidentelle et délibérée, violation de données, etc.
  - **Certaines conséquences de risques, uniquement sur les personnes** : santé et droits fondamentaux

- Le périmètre du projet est encore plus large
  - **Tous les produits et les services qui reposent sur des systèmes d'IA**
  - **Toutes les causes de risques**
  - **Toutes les conséquences des risques** : sur les personnes (matérielles, physiques et psychologiques), sur les organisations, sur l'environnement
- Et il est extensible: responsabilité sociétale, sécurité économique, etc. !