National Nuclear Security Administration

United States
Department of Energy
National Nuclear Security Administration

International Nuclear Security

Intelligence artificielle responsable pour l'atténuation des menaces internes

Jessica Baweja, Jon Barr,
Chantell Murphy

23 octobre 2025

PNNL-SA-216591

INS International Nuclear Security
Reducing Risk of Nuclear Terrorism

Principes fondamentaux de l'IA responsable dans le cadre de l'ITM

VALIDE ET FIABLE

Le système produit des résultats précis de manière constante, qui peuvent être utilisés en toute confiance pour l'usage auquel ils sont destinés.

SÛR

Le système fonctionne sans générer de risque pour les personnes, les biens ou l'environnement.

SÉCURISÉ ET RÉSILIENT

Le système peut se protéger des attaques et continuer à fonctionner même si des problèmes surviennent.

REDEVABLE

On sait clairement qui est responsable des actions et des décisions du système.

EXPLICABLE ET INTERPRÉTABLE

Les utilisateurs peuvent comprendre comment et pourquoi le système prend ses décisions.

CONFIDENTIALITÉ RENFORCÉE

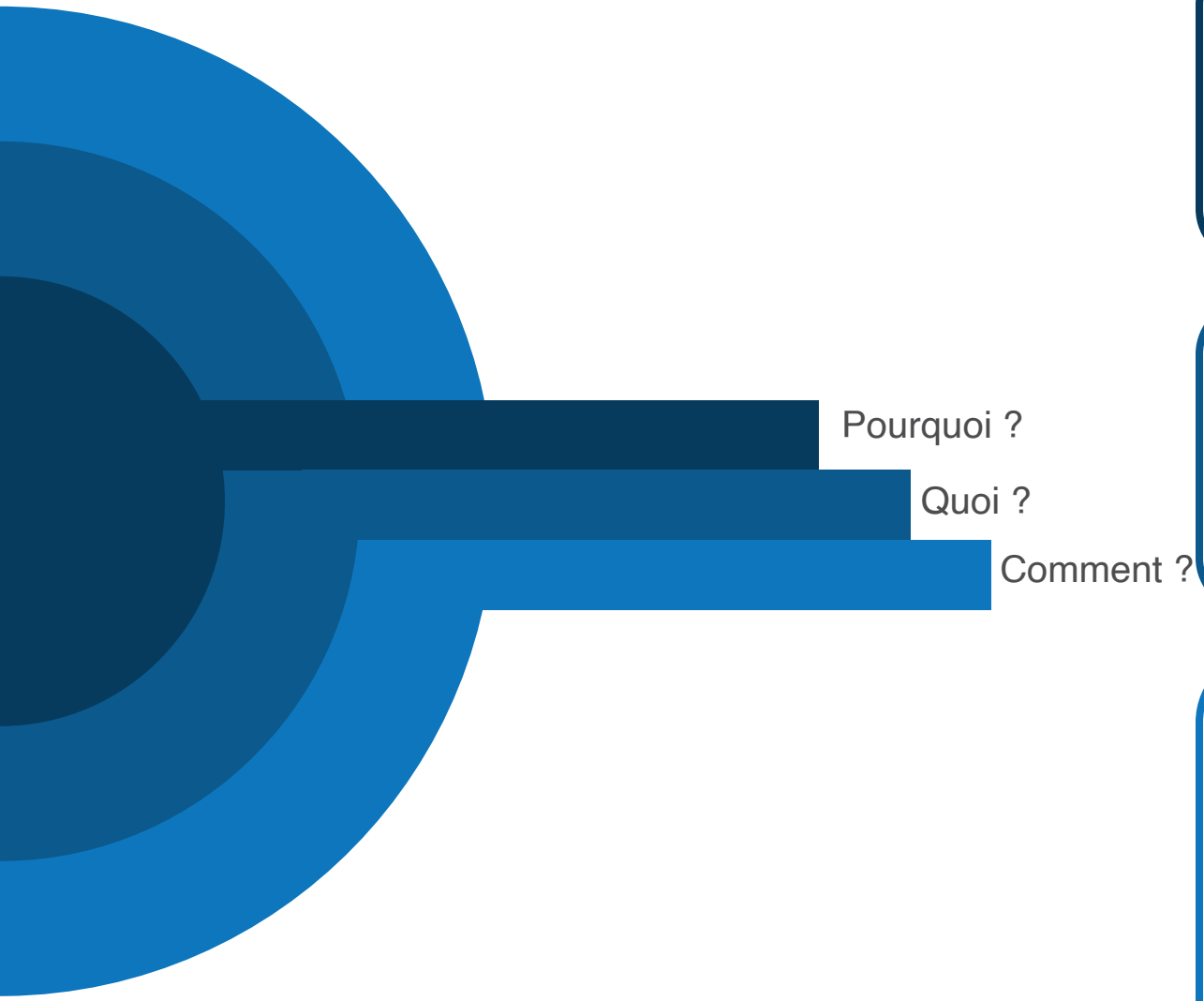
Le système protège les informations personnelles et respecte les droits au respect de la vie privée des personnes.

ÉQUITABLE

Le système traite toutes les personnes de manière égale et évite les biais préjudiciables.

Institut national américain des normes et technologies. (2023). Cadre de gestion des risques liés à l'intelligence artificielle (AI RMF 1.0).

Validité et fiabilité



POURQUOI c'est important :

- L'IA doit fonctionner de manière fiable dans toutes les conditions
- Les mauvaises performances créent des vulnérabilités sur le plan de la sécurité
- Différentes utilisations nécessitent différents niveaux de fiabilité

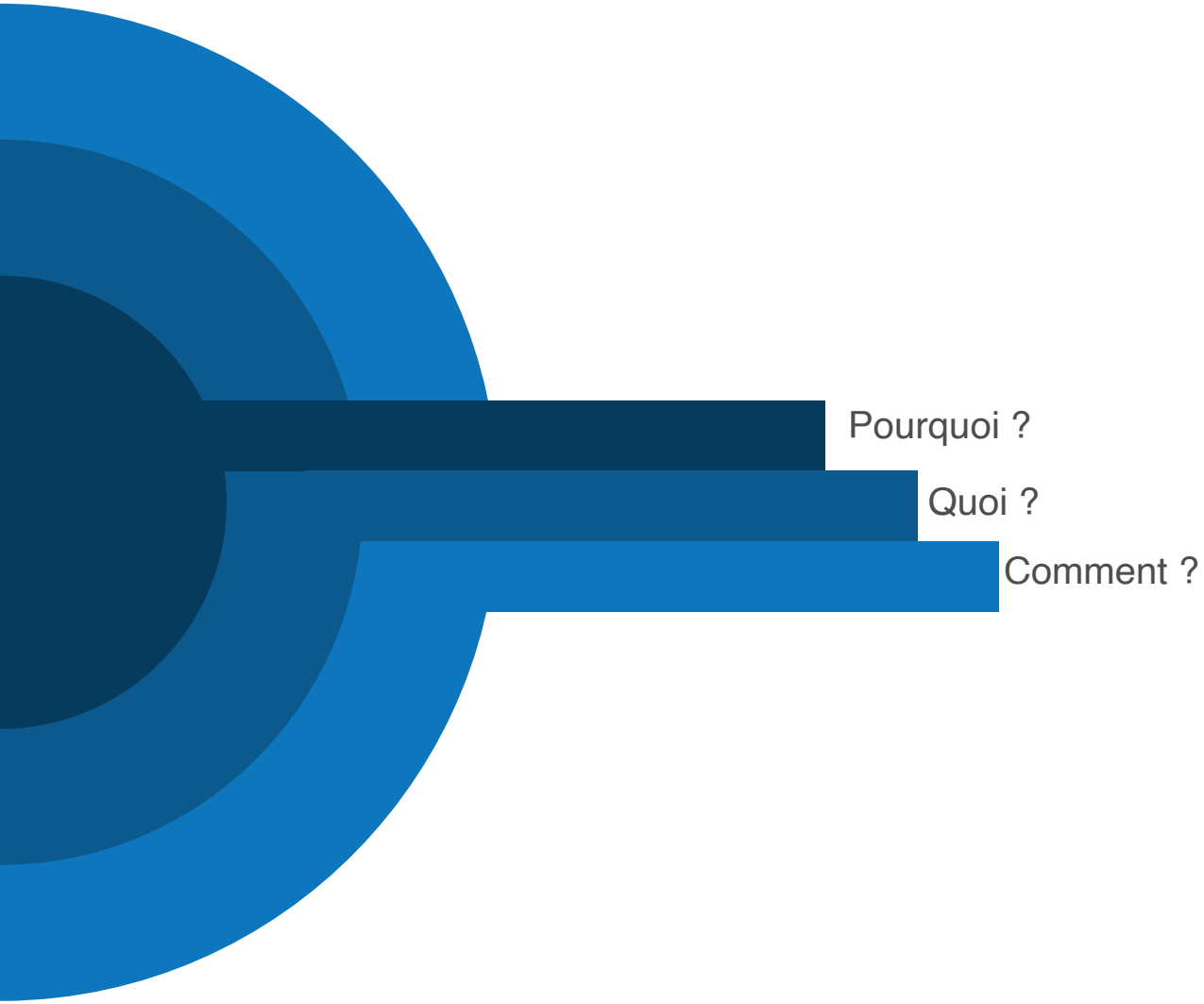
QUOI est concerné :

- S'assurer que les systèmes fonctionnent correctement dans tous les environnements
- Trouver l'équilibre entre un nombre trop élevé et un nombre trop faible d'alertes
- Vérifier régulièrement que les systèmes répondent aux exigences

COMMENT ceci est appliqué ?

- Tester les systèmes dans de nombreuses conditions différentes
- Vérifier régulièrement les performances avec des cas de test connus
- Ajuster les paramètres de sensibilité selon le besoin
- Suivre et examiner régulièrement les performances
- Définir des normes de performance minimales pour chaque application

Sûreté



POURQUOI c'est important :

- Une dépendance excessive à l'IA peut générer des failles de sécurité
- Les fausses alertes, ainsi que les menaces non détectées génèrent des problèmes
- Les systèmes automatisés peuvent créer des risques pour la sûreté

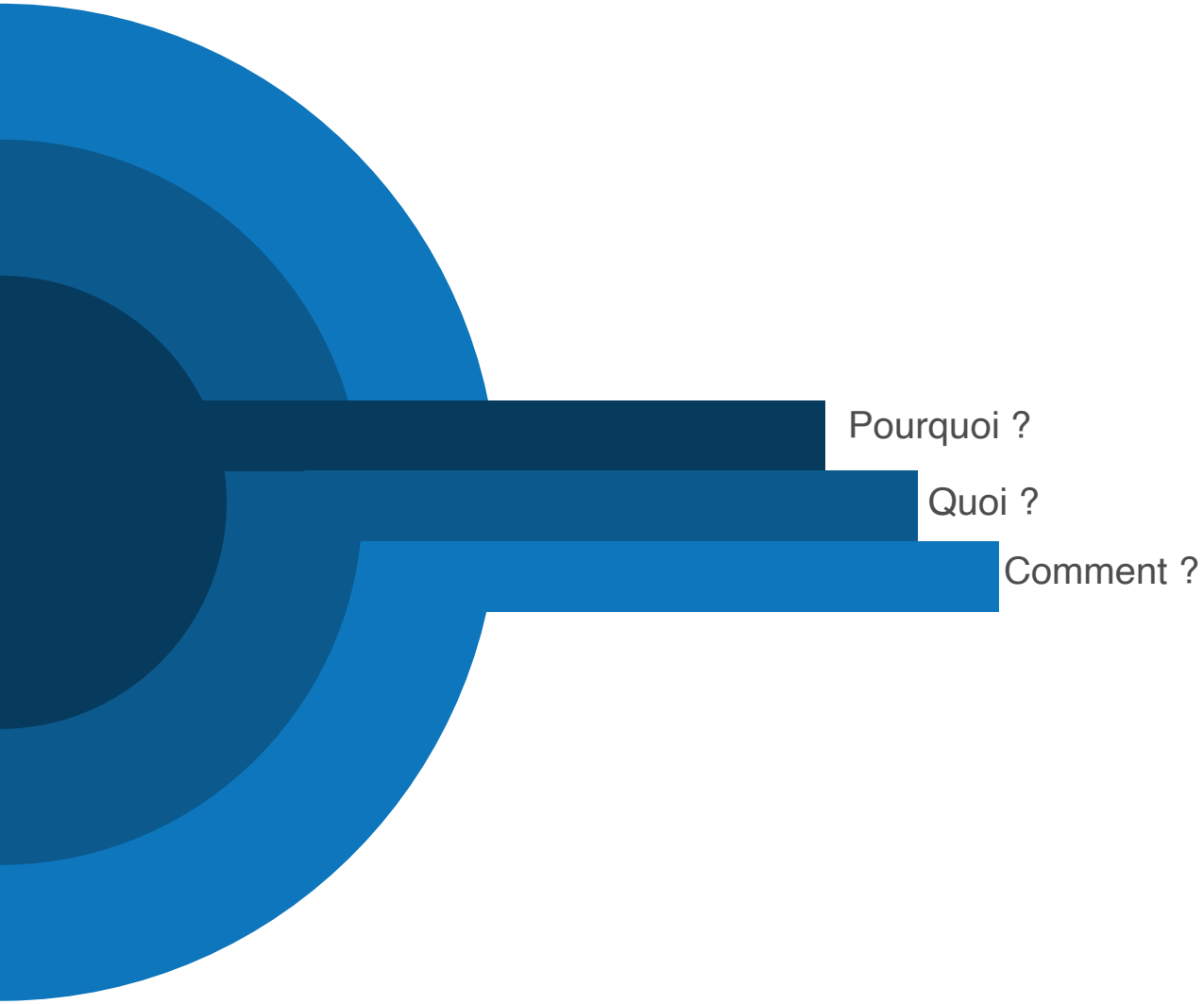
QUOI est concerné :

- S'assurer que l'IA améliore la sécurité plutôt que le contraire
- Maintenir des compétences humaines pointues en parallèle aux outils d'IA
- Prévenir les erreurs susceptibles d'affecter la sûreté

COMMENT ceci est appliqué ?

- Maintenir des méthodes de sécurité traditionnelles en parallèle à l'IA
- Élaborer des plans équilibrés pour intervenir quels que soient les risques
- Utiliser plusieurs mesures de sécurité qui se recoupent
- S'entraîner régulièrement aux tâches de sécurité sans IA
- Créer des procédures de contournement simples pour les systèmes automatisés

Sécurité



POURQUOI c'est important :

- Ces systèmes sont la cible d'attaques
- Ces systèmes pourraient faire l'objet d'abus et être utilisés pour surveiller certaines personnes de manière injuste
- L'accès à des données sensibles crée des risques d'utilisation abusive en interne

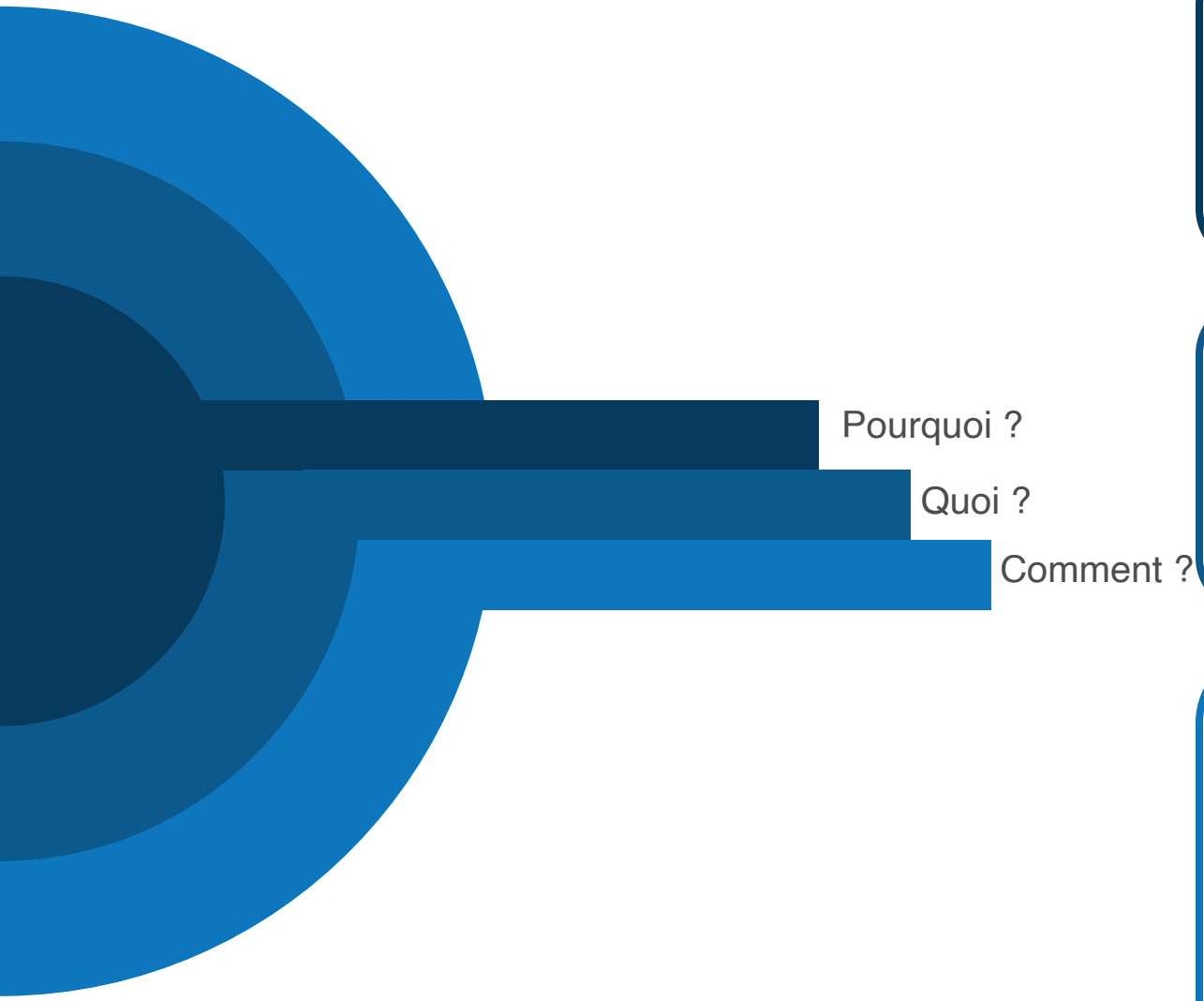
QUOI est concerné :

- Protéger les systèmes d'IA contre le piratage ou la manipulation
- Empêcher toute utilisation abusive par des personnes disposant d'un accès autorisé
- Maintenir une sécurité élevée pour le système d'IA lui-même

COMMENT ceci est appliqué ?

- Créer plusieurs couches de contrôle d'accès avec des journaux détaillés
- Mettre en place des garanties qui empêchent le ciblage d'individus spécifiques
- Tester régulièrement les contrôles de sécurité pour détecter leurs faiblesses
- Assurer une surveillance indépendante des schémas d'utilisation du système
- Créer des moyens de détecter et d'empêcher les utilisations abusives du système

Redevabilité



POURQUOI c'est important :

- Les systèmes d'IA aident à prendre des décisions importantes en matière de sécurité
- Sans responsabilité claire, personne « n'assume » de décisions
- Quand la responsabilité n'est pas clairement définie, des failles de sécurité apparaissent

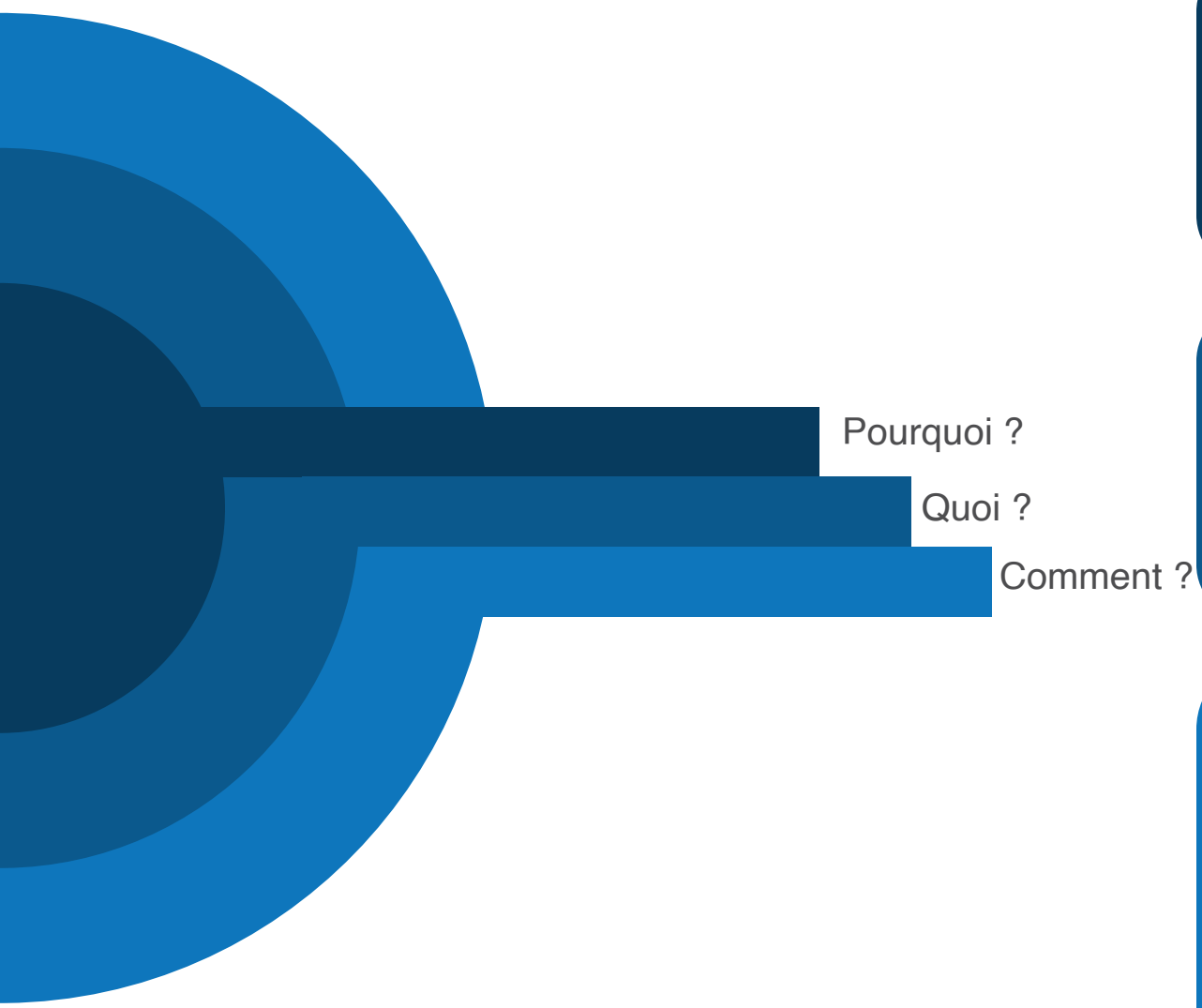
QUOI est concerné :

- Responsabilité claire relativement aux décisions prises par le système d'IA
- Rôles définis en matière de supervision humaine
- Processus de prise de décision transparent

COMMENT ceci est appliqué ?

- Établir des politiques claires indiquant qui examine les décisions prises par l'IA
- Configurer des cadres simples pour traiter les alertes du système
- Tenir des registres de toutes les décisions et de leurs auteurs
- Définir des étapes claires pour la remontée hiérarchique de différents types d'alertes
- Veiller à ce que les applications d'IA à haut risque soient toujours supervisées par des humains

Explicabilité



POURQUOI c'est important :

- Les systèmes d'IA aident à prendre des décisions importantes en matière de sécurité
- Sans responsabilité claire, personne « n'assume » de décisions
- Quand la responsabilité n'est pas clairement définie, des failles de sécurité apparaissent

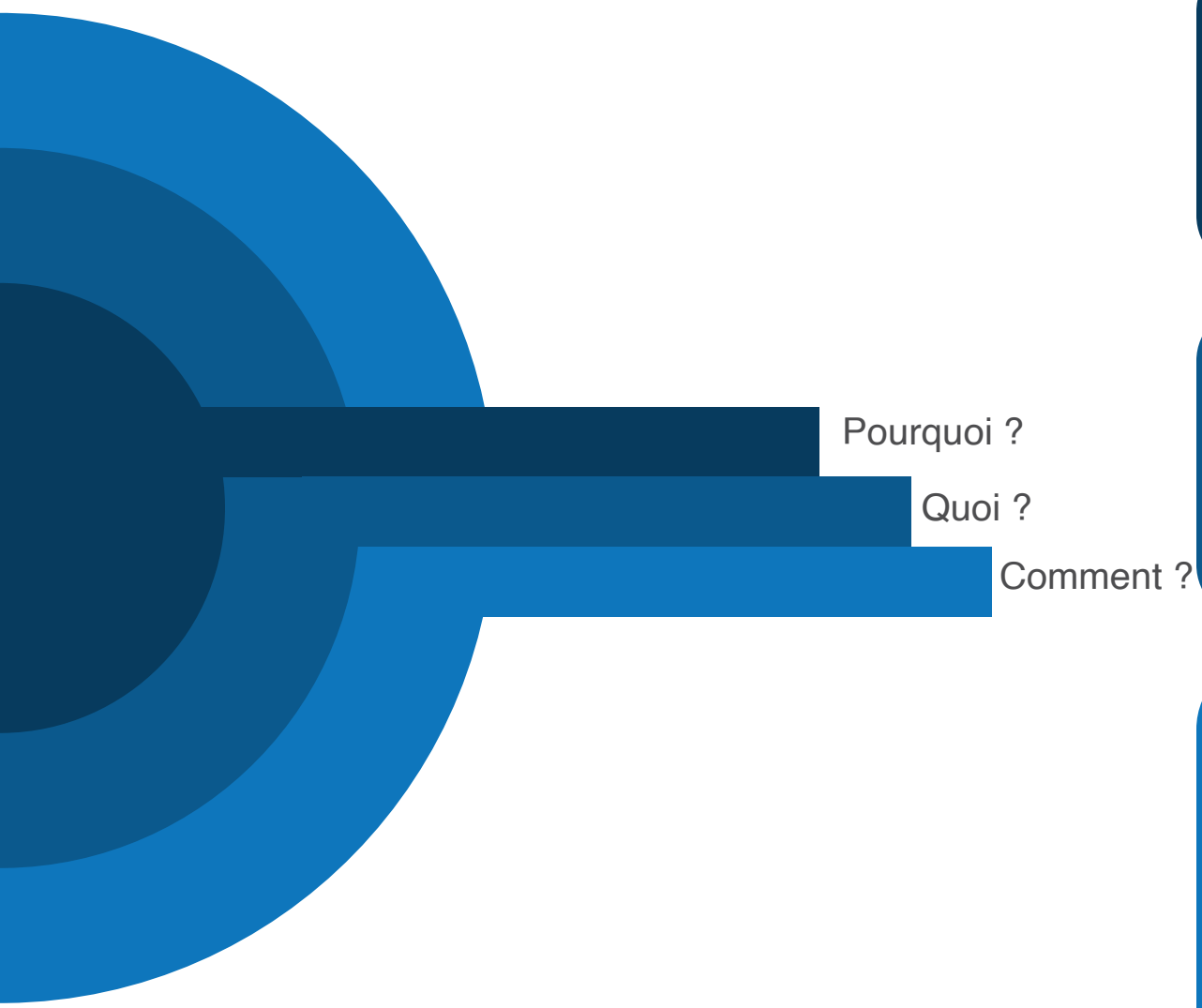
QUOI est concerné :

- Responsabilité claire relativement aux décisions prises par le système d'IA
- Rôles définis en matière de supervision humaine
- Processus de prise de décision transparent

COMMENT ceci est appliqué ?

- Établir des politiques claires indiquant qui examine les décisions prises par l'IA
- Configurer des cadres simples pour traiter les alertes du système
- Tenir des registres de toutes les décisions et de leurs auteurs
- Définir des étapes claires pour la remontée hiérarchique de différents types d'alertes
- Veiller à ce que les applications d'IA à haut risque soient toujours supervisées par des humains

Confidentialité



POURQUOI c'est important :

- Ces systèmes utilisent des informations très personnelles
- Les problèmes de confidentialité nuisent à la confiance et peuvent enfreindre la loi
- Les systèmes ont tendance à collecter plus de données que nécessaire au fil du temps

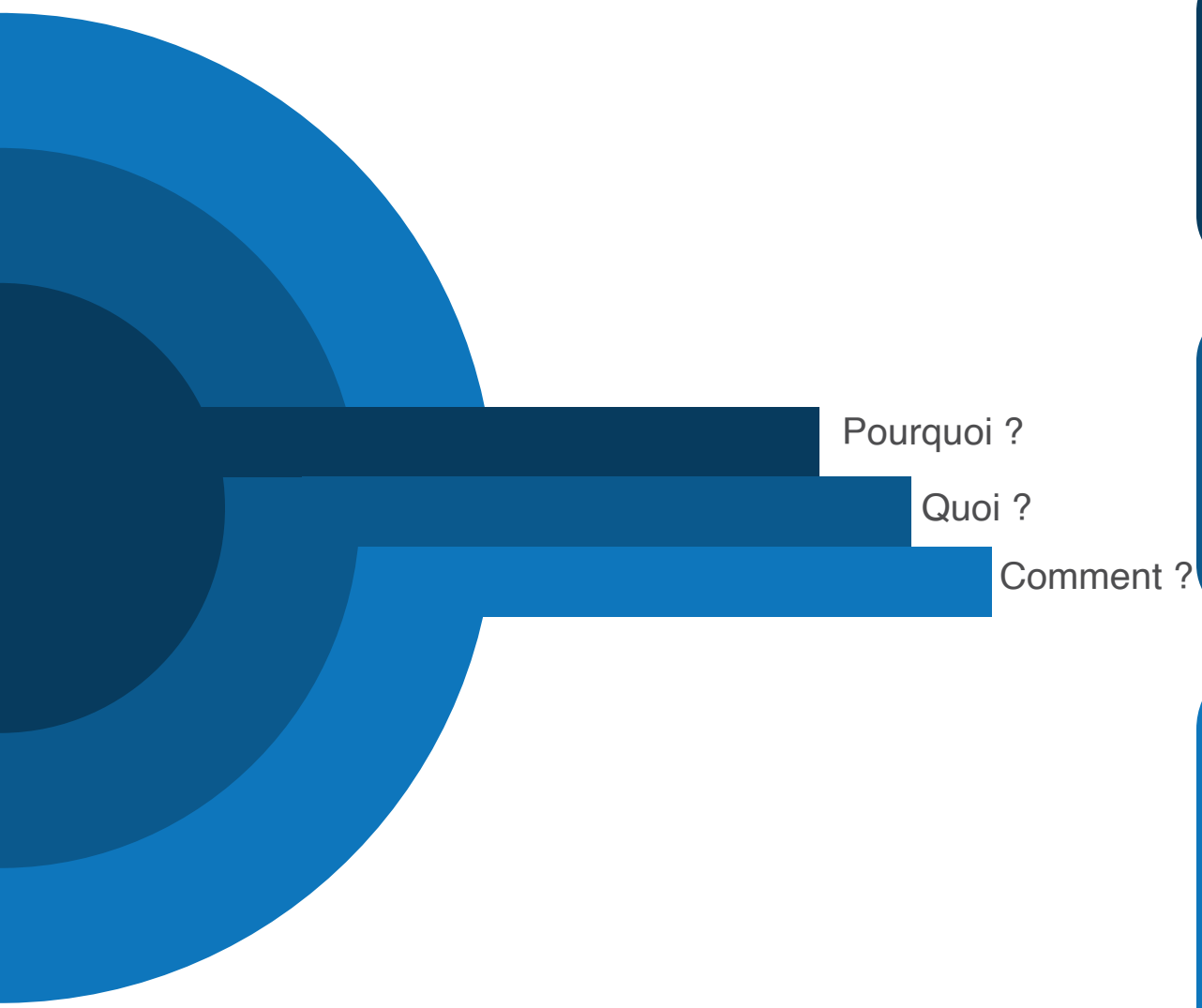
QUOI est concerné :

- Protéger les données personnelles contre l'accès non autorisé
- Équilibrer les besoins en matière de sécurité et le droit à la vie privée
- Limiter la collecte de données à ce qui est véritablement nécessaire

COMMENT ceci est appliqué ?

- Collecter uniquement les informations qui sont nécessaires
- Créer des contrôles stricts pour déterminer qui peut accéder aux données
- Définir des limites claires concernant l'utilisation des données
- Séparer la surveillance de la sécurité des évaluations professionnelles
- Élaborer et respecter des calendriers clairs pour la suppression des données

Équité



POURQUOI c'est important :

- Les systèmes d'IA peuvent fonctionner différemment selon les groupes
- Les données historiques contiennent souvent des biais
- Les systèmes injustes nuisent à la confiance et créent des problèmes de sécurité

QUOI est concerné :

- S'assurer que le système fonctionne bien pour tous les employés
- Empêcher les anciens préjugés d'influencer les nouvelles décisions
- Élaborer des processus équitables pour tous

COMMENT ceci est appliqué ?

- Tester les performances du système auprès de différents groupes avant de l'utiliser
- Impliquer des membres divers de l'équipe lors de l'examen des alertes
- Développer des normes claires qui tiennent compte des éventuels biais
- Employer plusieurs méthodes pour vérifier les informations
- Vérifier régulièrement si le système fonctionne aussi bien pour tout le monde

Écosystème de gestion des risques

- Responsabilité partagée en matière de gestion des risques liés à l'IA :
  - Responsables de la sécurité : définissent les seuils de risque et approuvent les politiques
  - Opérateurs système : évaluent les performances et l'efficacité quotidiennes
  - Équipes informatiques/de cybersécurité : sécurisent les systèmes et garantissent l'intégrité des données
  - Responsables de la conformité : vérifient la conformité réglementaire et les contrôles de confidentialité
  - Ressources humaines : répondent aux préoccupations du personnel et garantissent une application équitable
- Chacun a un rôle à jouer dans l'usage responsable de l'IA pour atténuer les menaces générées par les menaces internes.

Application responsable de l'IA

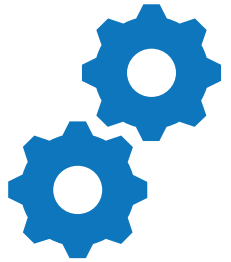
Conditions préalables

Politiques et procédures de sécurité claires

Intégration avec les systèmes de sécurité existants

Gouvernance des données définie

Formation du personnel de sécurité



Processus

Définir les objectifs de sécurité du système d'IA proposé

Évaluer la disponibilité et la qualité des données

Sélectionner les applications d'IA pertinentes

Appliquer avec des contrôles appropriés

Surveiller les performances et ajuster (amélioration continue !)

Synthèse et points clés à retenir

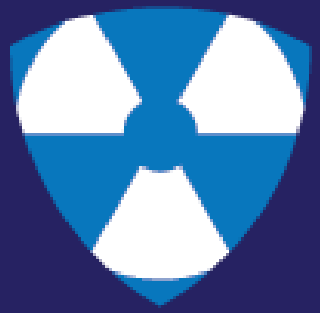
L'IA offre de puissantes capacités pour améliorer l'atténuation des menaces internes

Différentes applications entraînent des niveaux de risque et de complexité variés

Une application responsable nécessite d'équilibrer les avantages pour la sécurité avec les risques potentiels

La surveillance humaine demeure essentielle ; l'IA est un outil qui améliore les capacités humaines

Une approche fondée sur des principes garantit que l'IA renforce la sécurité tout en respectant les droits et les valeurs.



INS International Nuclear Security

Reducing Risk of Nuclear Terrorism