SYSTÈME D'IA

Définitions RIA – Article 3

1) **"système d'IA"** (SIA), un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels.

Définitions techniques

Tout système doté de **capacités d'inférence** est considéré comme un système d'IA (SIA). Cette capacité d'inférence, dans ce contexte, est identifiée par la concomitance de deux conditions (Cons. 12) :

- la présence d'un **processus consistant à générer des sorties** telles que des prédictions, du contenu, des recommandations ou des décisions, qui peuvent influencer l'environnement physique ou virtuel ;
- la capacité des systèmes d'IA à **inférer des modèles ou des algorithmes**, ou les deux, à partir d'entrées ou de données.

Cette formulation inclut les **IA basées sur l'apprentissage** (Machine Learning) et les **IA symboliques**.
La mise en conformité des SIA est un des éléments clés de l'*AI Act*.

Analyse du Hub France IA : Dans sa formulation actuelle, la définition inclut tous les systèmes dont le comportement résulte d'un processus d'apprentissage sur un ensemble de données initiales : régressions linéaires et logistiques, arbres de décision, réseaux de neurones, modèles LLM, etc. Les systèmes à base de connaissances ou de logique, par exemple les IA symboliques, sont également inclus.

Cette définition est indépendante des méthodes d'apprentissage utilisées. Les systèmes qui ont la possibilité d'évoluer en production (self learning capabilities), en utilisant les données les plus récentes collectées, sont également inclus dans cette catégorie. Toutefois, les modifications conceptuelles du modèle, du périmètre ou de l'utilisation qui n'étaient pas prévues à l'origine doivent être traitées comme un nouveau système (cf. modification substantielle).

Enfin, les systèmes d'IA sont conçus pour fonctionner à différents niveaux d'autonomie, ce qui signifie qu'ils bénéficient d'un certain degré d'indépendance dans leur action par rapport à une ingérence humaine et de capacités à fonctionner sans intervention humaine.

RISQUES

Définitions RIA – Article 3

2) **"risque"**, la combinaison de la probabilité d'un préjudice et de la sévérité de celui-ci.

Définitions techniques

Le *AI Act* prévoit 4 catégories pour classer les SIA en fonction de leur niveau de risque :

- **SIA interdits** (Art. 5) : l'utilisation de ces systèmes n'est pas autorisée, le risque associé à leur utilisation étant inacceptable ;
- **SIA à haut risque** (Art. 6) : il s'agit du niveau de risque pour lequel l'*AI Act* prévoit le plus d'exigences, en matière de documentation, de qualité des données, de performance, de sécurité, de mise en place d'un système de gestion et d'atténuation des risques et présence d'un contrôle humain dans le système.
- **SIA avec obligations particulières de transparence** (Art. 50) : exigences en matière de transparence pour l'utilisateur.
- **SIA à risque minime** : pas d'obligations.

Les deux premières catégories sont au cœur du présent règlement et sont définies par énumération, l'une à l'article 5, l'autre à l'article 6.1 et dans les Annexes I et III : une liste de cas entrant dans ces catégories a été établie et est susceptible d'être modifiée à l'avenir. Les deux dernières catégories ne sont pas explicitement définies par le règlement mais sont déduites de manière résiduelle.

Analyse du Hub France IA : La notion de risque doit être appréciée en fonction des impacts sur la santé, la sécurité, les droits fondamentaux des personnes (définis dans la Charte des droits fondamentaux de l'UE), ou d'autres aspects relatifs à la protection de l'intérêt public.

Sur les 445 occurrences du terme « risque » dans les articles de l'*AI Act*, 305 concernent l'expression "haut risque", 1 le risque de préjudice, 31 le risque systémique et 6 le risque au sens de l'article 79 "procédure applicable au niveau national aux systèmes d'IA présentant un risque".

Enfin, une définition de la manière de calculer le risque est considérée comme une condition préalable fondamentale au lancement d'un projet.

PERFORMANCES D'UN SYSTÈME D'IA

Définitions RIA – Article 3

18) **"performance d'un système d'IA",** la capacité d'un système d'IA à remplir sa destination.

Définitions techniques

Les performances d'un système d'IA sont généralement mesurées par différentes **métriques quantitatives**, qui sont utilisées pour évaluer la capacité d'un système à atteindre son objectif ou à résoudre la tâche qu'il s'est fixée. Ces mesures peuvent être fondées sur le jugement d'un expert pour répondre aux besoins spécifiques du système ou sur des mesures statistiques plus classiques que l'on peut trouver dans la littérature scientifique.
Ces métriques sont essentielles à la fois pour la **modélisation** et pour la **surveillance du système** en production. À la définition des métriques doit s'ajouter la sélection d'un benchmark qui doit être établi (littérature technique, besoin de l'entreprise, spécificité du cas d'usage, etc).

Analyse du Hub France IA : La sélection des métriques peut se faire selon les critères suivants :

- nature du problème (classification, régressions, etc.) ;
- type de données (structurées, non structurées) ;
- niveau d'interprétabilité requis ;
- exigences métier…

MODIFICATION SUBSTANTIELLE

Définitions RIA – Article 3

23) **"modification substantielle"**, une modification apportée à un système d'IA après sa mise sur le marché ou sa mise en service, qui n'est pas prévue ou planifiée dans l'évaluation initiale de la conformité réalisée par le fournisseur et qui a pour effet de nuire à la conformité de ce système aux exigences énoncées au chapitre III, section 2, ou qui entraîne une modification de la destination pour laquelle le système d'IA a été évalué.

Définitions techniques

La notion de « modification substantielle » est utilisée par l'*AI Act* en référence aux systèmes d'IA à haut risque (Cons. 177). Elle identifie chaque modification, non prévue lors de l'évaluation de la conformité **ou non attendue** à l'origine, appliquée après le déploiement d'un système d'IA et se décline en deux aspects principaux :

- **Changement de conception,** comme un changement dans l'architecture du système, un recalibrage du modèle ou un changement du processus de calibrage (prétraitement des données, feature engineering, processus d'optimisation, etc.), qui n'aurait pas été envisagé lors de l'évaluation de la conformité.
- **Changement de destination/utilisation/process,** comme l'évolution du périmètre d'application, l'application du système à différents cas d'usage, etc.

Ce changement peut être déclenché par une évolution imprévue du contexte ou une dégradation inattendue des performances du système.
Une modification substantielle doit être communiquée au régulateur et enregistrée dans la base de données de l'UE (Cons. 131). Enfin, **la soumission d'une nouvelle évaluation de la conformité pourrait être nécessaire** (Cons. 128).

Analyse du Hub France IA : Il est important de souligner que l'on peut s'attendre à des évolutions du modèle conformément à sa conception (par exemple, apprentissage en ligne ou apprentissage par renforcement). Étant donné que ces évolutions ont été prises en compte lors de la phase de conception et documentées (annexe IV, point 2.f), elles ne seront pas considérées comme des modifications substantielles.
Le règlement ne fournit pas de détails sur les cas relevant de cette définition. Selon l'interprétation du HFIA, une modification telle que le recalibrage d'un modèle prévu au moment d'une dégradation des performances ne serait pas considérée comme une modification substantielle, si elle avait été anticipée au moment du développement et des mesures correctives appropriées ont été définies dans le cadre du processus de surveillance.

En ce qui concerne les modèles d'IA à usage général (GPAIm), le HFIA s'interroge sur le fait de devoir considérer le fine-tuning comme une modification substantielle. La question reste à ce stade ouverte, et pourrait avoir un impact en faisant peser sur le déployer, les responsabilités du fournisseur (Art. 25 1.b).

Enfin, les événements enregistrés par le système (journalisation) doivent être utilisés pour évaluer la nécessité de déployer une modification substantielle.