HUB FRANCE IA

AI Act – Définitions clés

DONNEES BIOMETRIQUES

Définitions RIA – Article 3

34) **"données biométriques",** les données à caractère personnel résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, telles que des images faciales ou des données dactyloscopiques.

Définitions techniques

Les "données biométriques" incluent des données telles que le visage, les mouvements oculaires, la forme du corps, la voix, la prosodie, la démarche, la posture, le rythme cardiaque, la pression sanguine, l'odeur et la frappe au clavier. L'usage de données biométriques est une caractéristique sensible dans la détermination du niveau de risque d'un SIA.

Analyse du Hub France IA : Selon les considérants du RIA, les données biométriques peuvent permettre des fonctions d'authentification, d'identification ou de catégorisation des personnes physiques, ainsi que de reconnaissance de leurs émotions (Cons. 14[1]).

- Fonction d'identification biométrique : exploitation automatisée de caractéristiques physiques, physiologiques et comportementales (les données biométriques) d'une personne, aux fins d'établir l'identité d'une personne par comparaison des données biométriques de cette personne avec les données biométriques des personnes stockées dans une base de données de référence (comparaison 1-à-plusieurs – "Qui êtes-vous ?" - Cons. 15). L'utilisation de systèmes d'identification biométrique à distance en temps réel dans des espaces accessibles au public à des fins répressives **est interdite**, sauf si et dans la mesure où cette utilisation est strictement nécessaire eu égard à l'un des objectifs indiqués à l'Article 5 1.h de l'*AI Act* et autorisée par le droit de l'Union ou le droit national (Annexe III). Il convient de noter que la définition même de système d'identification biométrique à distance donnée dans l'*AI Act* ne met pas tant l'accent sur la distance que sur l'absence de **participation active** de la part du sujet.
- Fonction de vérification biométrique, ce qui inclut l'authentification (comparaison 1-à-1 - "Êtes-vous bien cette personne ?") : confirmation qu'une personne physique donnée est bien celle qu'elle prétend être dans le seul but d'avoir accès à un service, de déverrouiller un dispositif ou de disposer d'un accès sécurisé à des locaux. La fonction ne serait **pas interdite ni à haut risque**, considérant que la personne a donné son consentement pour être authentifiée (Cons. 15, Annexe III).
- Fonction de catégorisation biométrique, définie comme le classement de personnes physiques dans certaines catégories sur la base de leurs données biométriques. Ces catégories spécifiques peuvent concerner des aspects tels que le sexe, l'âge, la couleur des cheveux, la couleur des yeux, les tatouages, les traits liés au comportement ou à la personnalité, la langue, la religion, l'appartenance à une minorité nationale ou encore l'orientation sexuelle ou politique (Cons. 16). Cette fonction est classée comme **interdite** sauf dans le domaine répressif pour l'étiquetage ou le filtrage d'ensembles de données biométriques acquis légalement (Art .5 1.g). Dans ce cas, si le droit de l'Union ou le droit national l'autorise, la fonction est considérée comme à **haut risque** (Annexe III).
- Fonction de reconnaissance des émotions : un système d'IA permettant la reconnaissance ou la déduction des émotions ou des intentions de personnes physiques sur la base de leurs données biométriques. L'utilisation de ces systèmes d'IA pour inférer les émotions d'une personne physique sur le lieu de travail et dans les établissements d'enseignement **est interdite**, sauf lorsque l'utilisation du système d'IA est destinée à être mise en place ou mise sur le marché pour des raisons médicales ou de sécurité (Art .5 1.g); Dans les autres cas, cette fonction est considérée comme à **haut risque** (Annexe III).

[1] *Malgré le renvoi du considérant 14 de l'AI Act à la définition de données biométriques donnée par le RGPD (Art 4. 14), il est à noter que cette dernière intègre dans la notion la fonction de « identification unique » que ces données devraient permettre ou confirmer, ce qui n'est pas le cas dans le AI Act.*

MODELE D'IA A USAGE GENERAL

Définitions RIA – Article 3

63) **"modèle d'IA à usage général"**, un modèle d'IA, y compris lorsque ce modèle d'IA est entraîné à l'aide d'un grand nombre de données utilisant l'auto-supervision à grande échelle, qui présente une généralité significative et est capable d'exécuter de manière compétente un large éventail de tâches distinctes, indépendamment de la manière dont le modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval, à l'exception des modèles d'IA utilisés pour des activités de recherche, de développement ou de prototypage avant leur mise sur le marché.

Définitions techniques

L'*AI Act* prévoit des règles pour les modèles d'IA à usage général (GPAIm) et des règles supplémentaires plus spécifiques pour ceux qui présentent des risques systémiques. Ces règles doivent également s'appliquer lorsque ces modèles sont intégrés dans un système d'IA ou en font partie. Les obligations incombant aux **fournisseurs de ces modèles d'IA** doivent s'appliquer au plus tard lors de la mise sur le marché de ces modèles. Les fournisseurs de GPAIm doivent se conformer aux **obligations** suivantes (Art. 53):

- Établir et tenir à jour la documentation technique du modèle, y compris son processus d'entraînement et d'essai et les résultats de son évaluation. Les détails sont fournis à l'Annexe XI de l'*AI Act*.
- Respecter les lois européennes sur les droits d'auteur.
- Fournir les informations présentant les données utilisées pour entraîner le GPAIm. La granularité sera fournie par le Bureau de l'IA.
- Élaborer, tenir à jour et mettre à la disposition des **fournisseurs de systèmes d'IA**, qui ont l'intention d'intégrer le modèle d'IA à usage général dans leurs systèmes d'IA, la documentation leur permettant de bien comprendre les capacités et les limites du GPAIm. Les détails sont fournis à l'Annexe XII de l'*AI Act*.

Des exceptions à certaines exigences peuvent être faites pour les modèles open source, sauf s'ils présentent un risque systémique.
Ces obligations s'appliquent ici aux **fournisseurs de ces modèles d'IA** car ni les fournisseurs de systèmes d'IA ni les déployeurs ne sont en capacité d'avoir accès à ces informations.

Analyse du Hub France IA : La définition se fonde sur les principales caractéristiques fonctionnelles d'un modèle d'IA à usage général, en particulier la généralité et la capacité d'exécuter de manière compétente un large éventail de tâches distinctes. Ces modèles sont généralement entraînés avec de grandes quantités de données. Les modèles d'IA à usage général peuvent être mis sur le marché de différentes manières, peuvent être modifiés ou affinés et ainsi se transformer en nouveaux modèles. Ils ne constituent pas en eux-mêmes des systèmes d'IA mais ils en sont des composants essentiels incluant d'autres composants, tels qu'une interface utilisateur, pour devenir des systèmes d'IA (Cons. 97).
Quelques exemples connus de GPAIm sont les modèles BERT, GPT, Llama et Mistral. Ces modèles peuvent être utilisés tels quels dans des systèmes d'IA (ex: Copilot) ou intégrés dans des systèmes d'IA après fine-tuning, RAG, etc.
La régulation de ces modèles est apparue tardivement dans le processus d'élaboration de l'*AI Act*, ce qui explique **l'approche différenciée** pour la régulation de ces modèles par rapport aux SIA prédictifs. Cela est dû à la diffusion de cette technologie au grand public fin 2022.

HUB FRANCE IA

AI Act – Définitions clés

RISQUE SYSTEMIQUE

Définitions RIA – Article 3

65) **"risque systémique"**, un risque spécifique aux capacités à fort impact* des modèles d'IA à usage général, ayant une incidence significative sur le marché de l'Union en raison de leur portée ou d'effets négatifs réels ou raisonnablement prévisibles sur la santé publique, la sûreté, la sécurité publique, les droits fondamentaux ou la société dans son ensemble, pouvant être propagé à grande échelle tout au long de la chaîne de valeur.

*64) **"capacités à fort impact"**, des capacités égales ou supérieures aux capacités enregistrées dans les modèles d'IA à usage général les plus avancés.

Définitions techniques

Un modèle est présumé présenter un tel risque (à fort impact) lorsqu'il dépasse 10^{25} opérations en virgule flottante pendant la phase d'entraînement. Le nombre d'opérations utilisées dans le processus de formation est l'un des indicateurs, mais ce n'est pas le seul. La Commission pourra également déterminer qu'un modèle d'IA présente un risque systémique en se basant soit sur une alerte qualifiée du groupe scientifique ou sur la base d'autres critères, dont le nombre de paramètres utilisés, la taille de l'ensemble de données impliquées dans le processus de modélisation, le type de données d'entrée/sortie (texte, images, multimodal), l'impact sur le marché de l'Union et le nombre d'utilisateurs (Annexe XIII). Les critères et les seuils actuellement indiqués dans l'*AI Act* **pourraient évoluer** à l'avenir en fonction des progrès technologiques par actes délégués (Art. 51). Dès lors qu'un GPAIm remplit ces conditions, le fournisseur concerné doit en **informer la Commission** (Art. 52).

Les fournisseurs de modèles d'IA à usage général présentant des risques systémiques sont soumis, en plus des obligations prévues pour les fournisseurs de modèles d'IA à usage général, à **des exigences supplémentaires** d'évaluation, de transparence et de cybersécurité définies à l'Article 55. La conformité à ces exigences peut être établie grâce aux codes de bonnes pratiques de l'Article 56 ; elle est présumée en cas de respect de la norme européenne harmonisée qui reste à être adoptée.

Analyse du Hub France IA : Les GPAIm à risque systémique sont présumés **plus risqués** que les GPAIm classiques. En effet, les risques augmentent avec la taille (nombre de paramètres du modèle), la nature de l'ensemble de données, le nombre d'utilisateurs, etc. Par exemple, un ensemble de données plus grand contiendra potentiellement plus de données protégées, un modèle plus grand aura un impact environnemental plus élevé et plus d'utilisateurs peuvent avoir plus d'impacts négatifs.

Le Bureau de l'IA a lancé le 30 septembre les travaux de rédaction pour les **codes de bonnes pratiques**. Leur objectif est de faciliter l'application des dispositions de l'*AI Act* aux GPAIm en ce qui concerne notamment :
- La transparence et le droit d'auteur.
- Les règles de classification pour les GPAIm présentant un risque systémique.
- L'évaluation des risques et les mesures d'atténuation.

Le bureau de l'IA a annoncé qu'une version finale du texte est attendue pour **avril 2025**.

SYSTEME D'IA A USAGE GENERAL

Définitions RIA – Article 3

66) **"système d'IA à usage général"**, un système d'IA qui est fondé sur un modèle d'IA à usage général et qui a la capacité de répondre à diverses finalités, tant pour une utilisation directe que pour une intégration dans d'autres systèmes d'IA.

Définitions techniques

Un "système d'IA à usage général" est un modèle fondé sur un GPAIm. Comme tout SIA, il doit être qualifié au regard des trois catégories de risques identifiées par l'*AI Act*. En fonction des circonstances, les différents acteurs de la chaîne de valeurs de ces systèmes (par exemple, fournisseur du GPAIm et déployeur du système d'IA à usage général) doivent clarifier leurs rôles et responsabilités respectives pour permettre le respect des obligations applicables (Art. 25).

Les régimes de responsabilité avec ou sans faute seront complétés par les deux directives européennes attendues sur le sujet. A savoir, la directive européenne sur la responsabilité de l'IA proposée en septembre 2022 par la Commission européenne et dont le processus d'élaboration est toujours en cours et la révision de la directive 85/374/CEE sur les produits défectueux (PLD).

Analyse du Hub France IA : Le fournisseur de systèmes d'IA à usage général devra être très vigilant sur la **répartition des responsabilités** entre lui et le fournisseur du GPAIm.

Quelques exemples sont Chat-GPT, Dall-E et Gemini, et des systèmes d'IA à usage général développés par les entreprises.