AI Act – Opérationnalisation

MODÉLISATION (1)

Guider l'alignement entre les bonnes pratiques du processus de modélisation et les exigences du RIA, qui dépendent de la qualification du Système d'IA en cours de réalisation. Le niveau de détail et les exigences décrits dans ce livrable seront étroitement liés avec les spécifications des normes harmonisées à paraître dans le cadre du RIA.

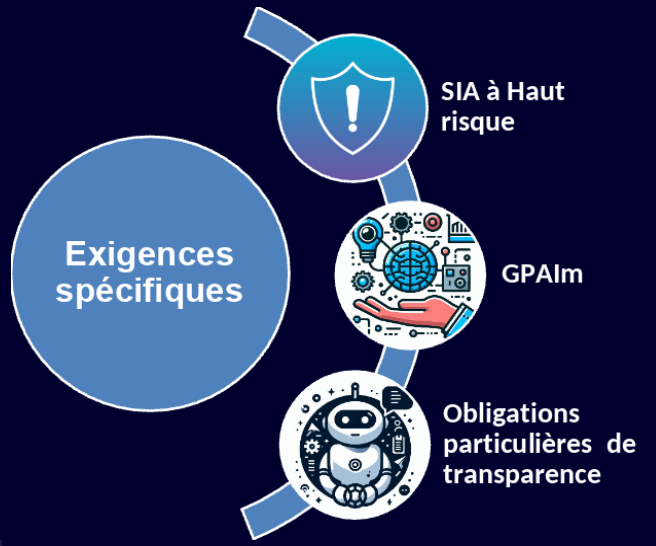
Étape 1 **Construction du modèle**	**Développer un système d'IA en suivant une méthodologie standardisée pour assurer cohérence et qualité.**
Étape 2 **Évaluation du modèle**	**Comparer les performances du modèle avec des processus existants ou des références humaines.**
Étape 3 **Sélection du modèle**	**Choisir le modèle présentant le profil le plus adapté vis à vis de critères relatifs aux exigences techniques, réglementaires et métier, en tenant compte de l'interprétabilité, les biais et les contraintes opérationnelles avant validation finale.**

(1) La phase et les étapes font référence au livrable « Opérationnaliser la gestion des risques » du *GT Banque* du Hub France IA.

HUB FRANCE .IA

AI Act – Opérationnalisation

MODÉLISATION



Les exigences et actions attendues par l'*AI Act* au cours du processus de modélisation varient en fonction du type de système d'IA : **systèmes d'IA (SIA) à haut risque**, **SIA à usage général** (systémique ou pas), **SIA avec obligations particuliers de transparence**.

- Ces catégories ne s'excluent pas mutuellement; un système peut donc potentiellement répondre aux exigences de ces trois catégories (cf. feuillet *AI Act différents scénarii de classification*).
- Pour plus d'informations sur le processus de modélisation, nous vous recommandons de consulter les lignes directrices élaborées par le *GT Banque.*

Préalables nécessaires

Normes internes transversales et réglementations spécifiques

Documents associés au projet (phases préalables)

Boîte à outils

- **Standards NIST** : Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations
- **MITRE ATLAS** - Base de tactiques et de techniques adverses contre les systèmes basés sur l'IA.
- **HELM** : reproducible and transparent framework for evaluating foundation models
- **OWASP** : LLM & Generative AI Security Risk
- **ANSSI** : Développer la confiance dans l'IA par une approche par les risques cyber. 7 février 2025
- **OECD:** Toolbox Trustworthy AI
- **À paraître** : Normes harmonisées validées par la Commission EU.

Notre check-list pour vous aider à opérationnaliser l'AI Act

- Les **réglementations spécifiques** à un domaine qui doivent être respectées en complément des exigences identifiées par le RIA.
- **Politiques de gestion de risques** à élaborer, et à mettre à jour (Art. 9).
- **Politiques de construction, évaluation,** (dont protocole de Stress Test), **sélection du modèle,** y compris GPAIm (Système de gestion de qualité - Art. 17 point b./c./d.)
- **Template documentation modèle** (Art. 11).
- **Politiques de gestion du risque IT** : mesures visant à assurer la **robustesse, la résilience et la cybersécurité** des systèmes (Art. 15).

- **Fiche projet,** transversale à toutes les étapes du projet et essentielle pour déterminer la diligence en matière de modélisation en fonction de la classification RIA du cas d'utilisation proposé.
- **Documentation technique** contenant les informations sur les données (train/validation/test), le processus de collecte et le prétraitement appliqué.

(1) L'explication plus détaillée de la Fiche Projet est disponible dans les définitions clés livrées par le GT BAIA.

HUB FRANCE IA

AI Act – Opérationnalisation

MODÉLISATION - FOCUS HAUT RISQUE

1. Construction du modèle

- **Gérer les risques.** Assurer l'alignement avec le système de gestion des risques de l'entreprise (Art. 9) et commencer à définir un système de log (Art. 12) capable d'identifier les situations potentiellement à risque, les modifications substantielles, les KPI de suivi (spécificités pour SIA visés à l'Annexe III, point 1.a).
- Dans le cadre du design du modèle, **sélectionner, prétraiter et enrichir les données** Focus sur les traitements de données propres à l'étape de modélisation (cf. Phase de gouvernance des données).
- **Mettre à jour la documentation technique (Annexe IV).**
- **Production de la notice d'utilisation** à destination des déployeurs.
- **Initier la conception d'un système permettant le contrôle du SIA par des personnes physiques,** en fonction du niveau de risque associé au SIA et de son cadre d'application (Art. 14).
- **Élaborer un système résilient** visant à atteindre un niveau approprié de précision, de robustesse et de cybersécurité, en s'appuyant sur des critères de référence et des méthodes de mesure appropriées (Art. 15, en attente des normes harmonisées).

Boîte à outils

- Système de gestion des risques
- Standard de construction des modèles candidats (stratégies d'apprentissage, architecture, familles de modèles compatibles avec les contraintes/obligations, etc)
- Template de documentation du modèle
- Template de construction de pipeline
- Outils de versioning
- Les réglementations sectorielles et/ou spécifiques en complément des exigences identifiées par le RIA.

2. Évaluation du modèle

Itérer jusqu'à converger vers les exigences attendues dans le respect des contraintes fonctionnelles, techniques et réglementaires :
1) Construire avec les différentes parties prenantes le protocole d'évaluation
2) Exécuter le protocole d'évaluation

Boîte à outils

- Protocole d'évaluation/cadre de test
- Métriques métiers et Data Science
- Outils de *benchmarking*
- Outils de détection de biais
- Outil(s) d'IHM, de contrôle humain

3. Sélection du modèle

- **Construction d'un protocole de sélection de modèle** avec les différentes parties-prenantes et déploiement

Boîte à outils

- Métriques métiers et Data Science
- Outils d'explicabilité et quantification d'incertitude

Notre *check-list* pour vous aider à opérationnaliser l'AI Act

- ❑ *Mise à jour des* **Fiches sur les jeux de données (d'entraînement, validation, test) utilisés, présentant le processus de prétraitement (art. 11, annexe IV, §2)**
- ❑ *Mise à jour de la* **Documentation technique du modèle**
- ❑ **Stratégie de gestion des logs**
- ❑ **Processus en place pour le déploiement du contrôle humain**
- ❑ *Mise à jour de la* **Notice d'utilisation pour les déployeurs**

❑ **Intégration Documentation technique**
1) Protocole d'évaluation : ensemble des expérimentations à conduire pour atteindre les exigences attendues et respecter les contraintes (précise l'interaction et intervention humaine)
2) Résultats du protocole d'évaluation : tableau de suivi des métriques et des biais par expérimentation

❑ **Intégration Documentation technique**
Procédures de validation et d'essai utilisées, y compris les informations sur les données (Annexe 4 2.g)) **– documenter le choix du modèle sur la fiche projet**

IHM : Interface Homme-Machine

AI Act – Opérationnalisation

MODÉLISATION – FOCUS GPAIm

1. Construction du modèle

Obligations incombant aux fournisseurs de modèles d'IA à usage général (Art. 53)*

- Établir et tenir à jour la documentation technique du modèle.
- Élaborer, tenir à jour et mettre à disposition la documentation permettant de bien comprendre les capacités et les limites du GPAIm.
- Assurer le respect des lois européennes sur les droits d'auteur.
- Fournir le résumé détaillé de la donnée d'entraînement du GPAIm.

Les obligations énoncées ne s'appliquent pas aux modèles d'IA qui sont publiés en licence libre et ouverte sauf s'ils présentent un risque systémique (Art. 53.2)

*Étendues aux entreprises qui *fine-tune* un GPAIm, exclusivement sur les parties modifiées/ nouvellement introduites (Cons. 109)

Boîte à outils

- Standard de construction du modèle
- Template de documentation du modèle
- Modèle de résumé des données d'entrainement du GPAIm (fourni par le Bureau de l'IA)
- Template de la déclaration de respect des lois européennes sur les droits d'auteur.

2. Évaluation du modèle

- **Classifier** des systèmes présentant **un risque systémique** (Art. 51), informer la Commission et/ou présenter les éléments pour demander une exemption (si applicable – Art. 52).
- Obligations complémentaires : évaluation du modèle et piste d'audit des tests.

Boîte à outils

- **NA**

3. Sélection du modèle

- **Optionnel :** v. le code de conduite interne si défini et applicable.

Boîte à outils

- **NA**

Notre *check-list* pour vous aider à opérationnaliser l'AI Act

- ❑ **Documentation technique du GPAIm,** y compris la description des limites et des faiblesses
- ❑ **Description des données** utilisées pour entrainer le GPAIm.
- ❑ **Déclaration** de respect des lois européennes sur les droits d'auteur.
- ❑ **Notification à la Commission européenne** (GPAIm à risque systémique)
- ❑ Définition d'un **mandataire** établi dans l'UE (si applicable)

- ❑ NA

- ❑ NA