# Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier

## Liste des questions soumises à commentaires

Ce document reprend la liste des questions soumises à commentaires du document de réflexion publié par le Pôle Fintech-Innovation de l'ACPR sur la gouvernance des algorithmes d'IA dans le secteur financier.

La consultation publique vise à recueillir l'avis des acteurs financiers et autres parties concernées par le sujet (chercheurs, prestataires, autorités de contrôle, etc.) sur les pistes de recommandations esquissées mais aussi, plus largement, tout commentaire utile, y compris sur l'adaptation des bonnes pratiques du superviseur.

Les réponses sont à envoyer à fintech-innovation@acpr.banque-france.fr avant le **4 septembre 2020**. Elles peuvent être illustrées par les cas d'usage de l'IA, en particulier du ML, en vigueur ou envisagés.

## Contexte

Cette partie de la consultation concerne le répondant, qui est par ailleurs invité à fournir quelques informations le concernant.

**QUESTION 1 : EXPÉRIENCE EN ML**

- Quelle est la nature de votre connaissance ou expérience de l'IA en général, et du ML en particulier (recherche, Data Science, connaissance opérationnelle, etc.) ?
- Si vous répondez au nom d'une entreprise du secteur financier, quel est le niveau de familiarité et d'expertise de vos équipes, aussi bien techniques que métiers, avec l'IA ?

**QUESTION 2 : MISE EN ŒUVRE DU ML *(QUESTION DESTINÉE AUX ENTREPRISES DU SECTEUR FINANCIER)***

- Quels algorithmes de ML sont en place dans votre organisation ?
- Pour chaque type d'algorithme, préciser les cas d'usage où ils sont utilisés, et le type d'environnement (développement, pré-production, production) ?
- Pour chaque cas d'usage, sur quels critères et quelle méthode d'évaluation a été fait le choix de l'algorithme retenu (performance pure, compromis explicabilité/efficacité, etc.) ?
- Quels sont les rôles respectifs des différentes équipes dans la conception et la mise en œuvre d'algorithmes de ML dans votre organisation (experts techniques, maîtrise d'ouvrage informatique traditionnelle, experts métiers, responsables conformité, etc.) ?

## Principe d'explicabilité

Cette partie de la consultation concerne les pages 12 à 18 du document de réflexion.

Ces pages ont en effet dégagé des exigences d'explicabilité sur la base des trois thèmes explorés, ainsi que sur celle d'une étude plus générale en matière d'IA – tant dans le secteur financier que plus généralement sur l'état de l'art de la recherche et dans les domaines les plus pertinents.

Plusieurs points restent à confirmer quant à la pertinence de ces exigences, qui forment l'objet de cette partie de la consultation.

**QUESTION 3 : DÉFINITION DES NIVEAUX D'EXPLICATION**

Les quatre niveaux d'explication ressortant de cette analyse (1 : observation, 2 : justification, 3 : approximation, 4 : réplication) sont-ils clairement définis ? Si non, préciser les motifs d'incompréhension.

**QUESTION 4 : ADÉQUATION DES NIVEAUX D'EXPLICATION**

Ces niveaux d'explication constituent-ils un ensemble de niveaux adéquat aux sens suivants :

- Ces niveaux couvrent-ils selon vous le spectre des applications actuelles ou futures de l'IA en finance, depuis une transparence totale (et donc une auditabilité facilitée) jusqu'à des algorithmes fonctionnant en boîte noire ?
- Le choix de quatre niveaux semble-t-il approprié (si non, en faudrait-il plus ou moins) ?

**QUESTION 5 : EXEMPLES PRATIQUES DE NIVEAUX D'EXPLICATION**

Le tableau présenté dans la section « Exemples de niveaux d'explication par cas d'usage » du document de réflexion donne, pour quelques cas d'usage de l'IA dans le secteur financier, une suggestion de niveau d'explication adapté.

- Ces suggestions vous paraissent-elles adaptées? Si non, pour quelle raison ?
- Sont-elles adaptées à vos propres scénarios d'utilisation de l'IA (préciser ces scénarios) ? Si non, dans quel sens ?

## Principe de performance

Cette partie de la consultation concerne les pages 10 et 11 du document de réflexion.

**QUESTION 6 : MESURES TECHNIQUES DE LA PERFORMANCE**

Quels commentaires suscitent de votre part les métriques techniques de la performance couramment utilisées (AUC ou score F1, score GINI, etc.), et notamment :

- Leur adéquation aux différents type d'algorithme d'IA ?
- Les méthodes adéquate de choix/sélection de ces métriques ?
- Les usages qui en sont faits (validation du modèle, choix du point de fonctionnement, détection de dérive, etc.) ?

**QUESTION 7 : MESURES FONCTIONNELLES DE LA PERFORMANCE**

- Quelles métriques fonctionnelles (ou « KPI ») vous paraissent-elles pertinentes ? Prennent-elles en comptes les aspects de conformité spécifiques aux processus concernés ?
- Par qui ces métriques devraient-elles être définies (équipes techniques, experts métiers, avec ou sans interaction avec les équipes conformité ou gestion des risques, …) ?

## Principe de stabilité

Cette partie de la consultation concerne les pages 11 et 12 du document de réflexion.

**QUESTION 8 : DÉRIVE TEMPORELLE DES MODÈLES**

- Quels risques induit, selon vous, la dérive temporelle des modèles de ML ?
- Quelles sont les méthodes utilisées pour y remédier ou du moins les circonscrire (*out-of-time testing*, déclenchement d'alertes en cas de dérive, etc.) ?

**QUESTION 9 : GÉNÉRALISATION DES MODÈLES**

- Quelles sont les limites identifiées au pouvoir de généralisation des modèles de ML, qu'elles soient liées au sur-apprentissage (*overfitting*) ou à des limites inhérentes au modèle ?
- Comment peuvent-elles être traitées (*out-of-sample testing*, etc.) ?

**QUESTION 10 : INSTABILITÉ DUE AU RÉENTRAÎNEMENT**

- Les phases de réentraînement (en mode périodique ou continu) sont-elles une source d'instabilité des modèles, selon votre expérience ?
- Quelles sont les techniques utilisées pour la limiter (jeux de données ou tests de non-régression, etc.) ?

## Principe de traitement adéquat des données

Cette partie de la consultation concerne les pages 7 à 10 du document de réflexion.

**QUESTION 11 : CONFORMITÉ RÉGLEMENTAIRE EN MATIÈRE DE DONNÉES**

Quelles méthodes ou techniques vous paraissent-elles indiquées pour s'assurer du respect des différentes contraintes réglementaires, notamment relatives au traitement des données :

- RGPD ?
- autres réglementations transverses ?
- réglementations sectorielles telles que la DDA ?

Préciser à quelle(s) phase(s) du processus (conception/apprentissage/prédiction) interviennent ces méthodes et techniques.

**QUESTION 12 : DÉTECTION ET REMÉDIATION DES BIAIS**

Quelles méthodes vous paraissent-elles indiquées pour analyser les biais, respectivement :

- dans les données sources utilisées par vos modèles de ML ?
- dans les algorithmes eux-mêmes ?
- dans les modèles produits par ces algorithmes et les décisions qu'ils engendrent ?

Plus précisément, quelles métriques d'équité (*fairness*) permettent d'identifier les biais, par exemple ceux à caractère discriminatoire ?

Quelles méthodes pour remédier aux biais non souhaités ainsi identifiés ?

## Intégration dans les processus

Cette partie de la consultation concerne les pages 21 à 24 du document de réflexion.

**QUESTION 13 : RÔLE DE L'IA**

- Quelles sont ou devraient être, selon vous, les grandes lignes d'une méthode d'analyse des composants d'IA selon leur intégration dans les processus métier ?
- Que devrait-elle permettre d'évaluer : criticité de leur fonction, caractère disruptif vis-à-vis du processus traditionnel, gain de productivité, types d'interactions humain/algorithme, etc. ?
- Que pensez-vous du maintien de processus « parallèles » confiés à des humains pour évaluer en continu (ou corriger) les résultats de l'algorithme ?

**QUESTION 14 : MÉTHODOLOGIE DE CONCEPTION DE L'IA**

- La méthodologie de conception des algorithmes d'IA devrait-elle différer des modèles traditionnels, et notamment des méthodes standards d'ingénierie logicielle ? Si oui, en quoi ?
- Comment faut-il, selon vous, que la chaîne de conception, de sélection et de validation des algorithmes de ML prenne en compte l'intégration de ces algorithmes aux processus métier?

## Contrôle interne

Cette partie de la consultation concerne les pages 24 à 28 du document de réflexion.

**QUESTION 15 : GESTION DES RISQUES**

- Quel est l'impact de l'introduction d'IA dans les différents processus métiers en matière de gestion des risques : présence de nouveaux risques associés ou amplification de risques existants (préciser de quelle nature : opérationnelle ou financière, juridique, etc.) ?
- Ces risques appellent-ils des mesures de gestion des risques spécifiques à l'IA (par exemple, calibrage des algorithmes de ML pour limiter l'exposition à tel ou tel risque) ?

**QUESTION 16 : VALIDATION FONCTIONNELLE**

- Quel doit être le processus de validation fonctionnelle avant mise en production ?
- La validation fonctionnelle devrait-elle être réitérée en cas de déploiement d'une version corrective ? Préciser si la réponse dépend du type de mise à jour.
- Comment doit être assuré le monitoring du ML en production sur le plan des risques métier ?

**QUESTION 17 : POLITIQUE DE CHANGEMENT DE MODÈLE *(MODÈLES DE RISQUES INTERNES)***

- Selon vous, à quelles conditions les algorithmes de ML peuvent-ils être utilisés pour les modèles Bâlois ou les modèles internes en assurance ?
- Comment peuvent-ils être pris en compte de façon adéquate dans la politique de changement de modèle de l'établissement concerné ?

**QUESTION 18 : VALIDATION TECHNIQUE**

- Quel doit être le processus de validation technique avant mise en production ?
- Comment le monitoring du ML peut-il être assuré en production sur le plan technique ?

## Sécurité et externalisation

Cette partie de la consultation concerne les pages 28 à 31 du document de réflexion.

**QUESTION 19 : EXTERNALISATION**

L'usage de l'IA introduit-t-il des enjeux ou des risques spécifiques lorsque son développement ou son exploitation sont externalisés ? Si oui, lesquels ?

**QUESTION 20 : SÉCURITÉ**

- Quel est l'impact de l'introduction du ML sur la sécurité des systèmes d'information (SSI) au sens classique ?
- Quels types d'attaques des modèles de ML vous semblent-ils les plus importantes, tant en termes de probabilité de survenue que d'impact en cas de succès (attaques causatives, attaques par substitution de modèle, attaques « adversarielles », etc.) ? Préciser selon le type de modèle de ML, selon le cas d'usage, et selon l'environnement (serveurs dédiés ou infrastructure externalisée sur le Cloud, etc.)

## Approche multifactorielle de l'évaluation

Cette partie de la consultation concerne les pages 32 à 35 du document de réflexion.

Ces pages ont en effet suggéré la mise en place d'une approche multifactorielle dans l'audit des processus mettant en œuvre de l'IA. Les questions suivantes permettent de préciser cette approche.

**QUESTION 21 : ÉVALUATION ANALYTIQUE**

- Quels éléments parmi les suivants sont, selon votre expérience, disponibles pour l'évaluation des algorithme d'IA : le code source ? la documentation ? les modèles résultants ? les données d'apprentissage et de validation ? Précisez si la réponse dépend du type d'algorithme, du cas d'usage étudié, ou du contexte de l'évaluation (validation interne, audit externe, etc.)
- Quelles méthodes (standards ou non) de documentation du ML sont-elles utilisées pour décrire les algorithmes, les modèles et/ou les jeux de données ?

**QUESTION 22 : ÉVALUATION EMPIRIQUE**

- Laquelle des deux approches de Benchmarking ou de mise en concurrence de modèles (cf. section « Évaluation empirique » du document de réflexion) vous semble-t-elle la plus appropriée ?
- La nature des architectures de Data Science au sein des organismes concernés est-elle suffisamment modulaire pour permettre ce genre de test fonctionnel au niveau des données ou de l'algorithme ?
- Le schéma de données est-il assez flexible pour se prêter à du *Benchmarking* sans faire reposer de contrainte d'intégration des données sur le superviseur ?
- De façon analogue, ce schéma est-il assez ouvert et documenté pour que le superviseur puisse mettre son propre modèle en concurrence sans pâtir d'une asymétrie informationnelle ?

**QUESTION 23 : MÉTHODES EXPLICATIVES**

- Quelles méthodes explicatives post-modélisation parmi celles décrites dans l'annexe « Recension des méthodes explicatives en IA » du document de réflexion sont actuellement en production parmi les divers cas d'usage de l'IA dont vous avez connaissance ?
- Avez-vous connaissance d'autres méthodes d'explication algorithmique que celles décrites ici ? Si oui, lesquelles ? Ont-elles déjà été mises en pratique ?
- Les méthodes explicatives utilisées varient-elles en fonction du type d'algorithme ?
- Varient-elles selon l'audience de l'explication et, si oui, comment ?
- Varient-elles selon le niveau de risque associé au processus et, si oui, comment ?