

Exigences réglementaires et éthiques des projets IA

par Julien Nguyen

Temps de lecture estimé : 8 minutes

Dans la continuité de notre [précédent article](#), nous proposons d'aborder les projets d'IA dans le secteur financier, un secteur des plus régulés qui soit. Aussi, il est impératif d'être attentif au respect des exigences réglementaires et éthiques qui impactent fortement les projets IA.

Exigence d'explicabilité¹ et gestion du risque modèle

Dans la banque-assurance, toute décision automatisée portant sur un client (*scoring* de crédit, tarification d'assurance, détection de fraude...) est scrutée par les régulateurs, notamment l'ACPR, qui attendent la justification de la décision prise ; ce qui peut être difficile quand on sait que les réseaux de neurones sur lesquels sont construits les modèles d'IA sont des « boîtes noires ». Il est aussi attendu la gestion du modèle de risque avec le processus de validation interne et l'auditabilité. **Si un modèle d'IA déployé n'est pas capable de fournir des justifications compréhensibles, il présente un risque de non-conformité aux principes de transparence.** Ainsi de nombreuses banques renoncent à utiliser des algorithmes trop complexes. Mieux vaut un modèle plus simple, un peu moins performant, mais explicable, qu'un « super algorithme » potentiellement non admissible par les régulateurs.

Comme le rappelle une note récente de l'ACPR consacrée à l'usage du *machine learning* dans les modèles internes IRB, **l'emploi d'algorithmes complexes demeure aujourd'hui « marginal » dans les banques européennes.** Les auteurs constatent pourtant que ces techniques pourraient « améliorer la qualité prédictive » des paramètres de risque et, pour certains

¹ Dans l'IA, l'explicabilité est la capacité de mettre en relation et de rendre compréhensible les éléments pris en compte par le système d'IA pour la production d'un résultat

portefeuilles, entraîner une « réduction des exigences de fonds propres ». Le frein majeur reste la faible interprétabilité : même les approches d'explicabilité ex-post n'offrent qu'une vision partielle du raisonnement du modèle, compliquant sa validation par le superviseur. La note souligne également que l'arrivée de CRR3/CRD 6 pourrait accentuer la variabilité des niveaux de capital si l'usage du machine learning (ML) n'est pas encadré par des recommandations techniques précises. Pour lever ces blocages, l'ACPR préconise de favoriser des modèles de *machine learning* « nativement interprétables », capables de concilier performance et transparence, tout en rappelant que la gouvernance des données et la montée en compétence des équipes demeurent des prérequis indispensables.

Protection des données et vie privée

Les lois de protection des données (RGPD en Europe) limitent fortement ce qui peut être fait, en particulier lorsqu'il s'agit de croiser les données ou de les utiliser pour de nouveaux usages. Un projet d'IA peut aiguiser sur un veto des juristes de l'entreprise, si, par exemple, il envisage d'envoyer des données client dans un cloud fournisseur externe.

Evitement des biais et discrimination

Les autorités de régulation financières et de protection des consommateurs veillent de près à ce que les algorithmes n'introduisent pas de discrimination injuste. Aux Etats-Unis, par exemple, la loi *Equal Credit Opportunity Act* (ECOA) s'applique aux modèles de *scoring*, **avec à la clé des sanctions si l'algorithme introduit de la discrimination**. A titre d'exemple, citons [l'Apple Card](#) qui a conduit le département des services financiers de l'État de New York (NYSDFS) à ouvrir une [enquête pour sexisme](#).

En Europe, la vigilance s'articule essentiellement sur l'article 22 du RGPD, qui donne à tout emprunteur le droit de ne pas être soumis à une décision entièrement automatisée ayant un impact juridique significatif et d'obtenir une intervention humaine ainsi qu'une explication intelligible. Le règlement européen sur l'IA (AI Act) classe explicitement les systèmes de « *credit scoring* » parmi les IA à haut risque ; il impose des tests de robustesse statistique, de la documentation des variables sensibles et des audits réguliers de non-discrimination. Côté prudentiel, les « **Guidelines on Loan Origination & Monitoring** » de l'Autorité européenne bancaire exigent désormais que les établissements identifient et corrigent les biais avant la mise en production d'un modèle, et qu'ils démontrent aux superviseurs que les résultats sont « équitablement répartis entre les catégories protégées ». En France, la CNIL et le Défenseur des droits ont appelé dès 2020 à une mobilisation générale contre les biais algorithmiques dans la finance, rappelant que toute discrimination fondée sur le sexe, l'origine ou le lieu de résidence reste pénalement sanctionnée. Des enquêtes récentes montrent que le phénomène n'est pas théorique : le podcast d'Algorithm Watch sur la « débanquarisation » a mis

en lumière des fermetures automatisées de comptes touchant disproportionnellement certaines minorités.

Aussi, beaucoup de projets d'IA sensibles intègrent désormais des phases de tests d'équité et d'audit éthique avant déploiement.

Evolution et incertitude réglementaire

Le cadre légal autour de l'IA se précise. Le règlement européen sur l'IA (AI Act) est entré en vigueur le 13 juin 2024 ; il est possible que certains modèles soient considérés par l'IA Act comme modèles à « haut risque » nécessitant des obligations de transparence, de documentation, voire de certification du système. Ce contexte incite les institutions à l'extrême prudence, et donc à freiner ou encadrer strictement certaines expérimentations.

Exemple

Une banque qui envisageait un *chatbot* conseiller financier a décidé de limiter drastiquement les types de conseils qu'il pourrait donner, de peur qu'un client ne suive une recommandation erronée et engage sa responsabilité. Le projet a été finalement abandonné en raison de la forte réduction de son champ fonctionnel.

Conservatisme et validation interne dans les projets d'IA : un enjeu stratégique

Les instances internes de contrôle – telles que la conformité, la gestion des risques ou les services juridiques – jouent un rôle décisif. Ces parties prenantes peuvent bloquer un projet lors des dernières revues de risque si elles ont été peu ou tardivement impliquées.

Les comités d'éthique des modèles et de conformité exigent une préparation rigoureuse : capacité à expliquer les modèles, démonstration de l'absence de biais, définition claire des indicateurs de contrôle. L'absence de cette préparation conduit fréquemment à des rejets, parfois à un stade très avancé. **Pour éviter ces échecs coûteux, de plus en plus d'organisations intègrent désormais les experts conformité dès la phase de conception.**

En synthèse, le contexte **réglementaire et éthique** agit comme un filtre sévère sur les projets d'IA dans la finance. L'acceptabilité devient un critère aussi important que la faisabilité. Les échecs dans ce domaine sont souvent liés à une sous-estimation initiale de ces exigences : vouloir aller trop vite sans intégrer la conformité « by design », ou ignorer les signaux faibles éthiques dans les données.

Il est recommandé de faire **auditer les algorithmes**, de mettre en place une **charte d'IA** responsable et d'**échanger avec les régulateurs** tout au long du projet.

Autres facteurs d'échec

Coûts sous-estimés et dépassements budgétaires

Le développement d'une IA peut s'avérer plus onéreux que prévu : besoin d'infrastructures de calcul important, nécessité d'acheter des données externes, qualification de la donnée, périodes d'entraînement, coût humain... les coûts grimpent vite. Ainsi un projet peut être stoppé lors d'une revue budgétaire si les dépenses dérapent sans résultats tangibles à court terme.

[Gartner](#) citait les coûts croissants comme l'un des motifs majeurs d'abandon de projets d'IA génératives récents.

ROI et valeur commerciale incertaine

Lié au coût, le manque de résultats mesurables clairs (chiffre d'affaires additionnel, économies, amélioration de la satisfaction...) condamne vite un projet. D'après Gartner toujours, **une valeur commerciale peu claire est l'un des premiers facteurs d'échec constatés sur les projets d'IA générative.**

Dépendance vis-à-vis d'un fournisseur ou d'une technologie

Certaines entreprises s'appuient sur des solutions externes (éditeurs de logiciels d'IA, fintechs spécialisées) pour accélérer leurs projets. Or, le fournisseur peut faire faillite, changer son offre ou ses prix, ou rencontrer un problème. **On a vu des fintechs IA très prometteuses disparaître faute de financement [en 2022](#) lors du retournement du marché des technologies**, laissant leurs clients sans support. De même miser sur une technologie propriétaire qui devient obsolète peut forcer à tout reprendre à zéro. Ces aléas externes font partie des facteurs d'échec rarement mis en avant, mais bien réels dans les retours d'expérience.

Pour conclure

Le secteur financier, conscient d'avoir "beaucoup à gagner mais aussi beaucoup à perdre" avec l'IA, ajuste progressivement son tir. En doublant les investissements dans l'IA d'ici 2027, les institutions financières montrent leur conviction que l'IA est un levier d'avenir – à condition d'intégrer les leçons du passé. Celles-ci pourraient se résumer ainsi : **recentrer l'IA sur la création de valeur réelle, construire des bases de données solides, impliquer étroitement l'humain** (clients, collaborateurs, régulateurs) **à chaque étape.**

Un adage semble émerger des nombreuses post-mortems de projets : "Commencez petit, pensez grand, échouez vite (si vous devez échouer) et apprenez toujours".