Réponse au document de réflexion de l'ACPR relatif à la gouvernance des algorithmes d'intelligence artificielle « IA » dans le secteur financier

PREAMBULE : Le Centre des Professions Financières, association à vocation éducative d'intérêt général dont l'objectif est de comprendre et faire comprendre les professions financières et leurs évolutions a réuni, sous l'égide du Club des marchés financiers, un groupe de travail afin de répondre au document de réflexion de l'ACPR.

Questions – 1 & 2 : Expérience en Machine Learning et Mise en œuvre des projets d'IA/ML

Pour répondre à cette question, les participants au groupe de travail, membres du Centre des professions financières, ont décrit les travaux d'IA auxquels ils participent et les cas d'usage concernés.

Une société d'assurance : le développement d'un robot advisor sur l'activité d'épargne collective et notamment un produit PER

Un éditeur de logiciels spécialisé dans des solutions clefs en main de gestion du risque de crédit : Automatisation des processus de gestion des risques et notamment le risque de crédit (octroi/scoring) avec une assistance virtuelle

Une société d'assurance – L'utilisation de la méthode NLP pour la classification documentaire et le développement d'un robot advisor dans le cadre de DDA (Directive sur la distribution des assurances).

Une société de services en informatique – Développement d'outil intégrant l'IA.

Une société de crédit à la consommation – mise en place de modèle des risques, notamment dans la lutte contre la fraude et les modèles d'octroi de crédit (modèle externe).

Une Université-Laboratoire de recherche – Lecture automatique de RIB – automatisation des process pour extraire des clauses et informations dans le cadre de projets AML

Un cabinet de conseil– L'automatisation du processus de veille réglementaire en utilisant le Machine Learning (NLP) pour analyser les textes réglementaires au niveau national et européen.

Questions 3 à 5 : Niveaux d'explication : *Les quatre niveaux d'explication ressortant de cette analyse (1 : observation, 2 : justification, 3 : approximation, 4 : réplication) sont-ils clairement définis et appropriés ?*

En pratique, l'auditabilité d'un algorithme nécessite à la fois une très grande maîtrise des méthodes employées **et une compréhension de son usage.**

Ces quatre niveaux nous semblent donc appropriés et complémentaires. Néanmoins, la réplication peut être difficile à réaliser.

Une problématique d'accès à l'hyperparamètre se pose néanmoins dans le cas des modèles en Open Source et les modalités d'explication des modèles en cas d'externalisation du processus sont plus compliquées à mettre en œuvre (notamment les niveaux 3 et 4) - (cf question 19).

Questions – 6 à 10 : Principe de performance

La performance de l'IA se fait au cas par cas et selon les modèles de ML utilisés :

- NLP (Natural Language processing): l'indicateur de performance est lié à l'output (type de document et informations ressorties par les algorithmes développés en IA) ; les indicateurs de performance sont plus faciles à définir.
- Autres méthodes de ML notamment en finance : la performance ne peut pas toujours être mesurée dans certains cas d'usage (comme les prédictions sur l'évolution des marchés)

Il faut être innovant sur les indicateurs et procéder à des générations de données synthétiques en utilisant une approche méthodologique avec des notions d'incertitude et une méthode d'active Learning.

La stabilité du modèle est essentielle et le recours à des experts métiers est indispensable pour prévenir et gérer les dérives des modèles.

La détection de dérives est l'une des mesures les plus difficiles à mettre en place, notamment lorsque l'algorithme est complexe, les modèles de machine learning en général n'étant pas capables de s'auto-recadrer tous seuls, il faut mettre en place un processus de détection des dérives des algorithmes (lorsque les résultats ne sont plus cohérents).

La surveillance des algorithmes doit être intégrée dans les dispositifs de contrôle permanent.

Il est aussi nécessaire de procéder à l'archivage des anciennes versions des modèles utilisées, notamment les plus structurantes (ce qui suppose une analyse d'écarts entre les différentes versions).

Question 11 : Principe de traitement adéquat des données

Il est nécessaire de bien respecter les dispositions des textes portant sur la protection des données (personnelles et non personnelles), notamment :

- Le règlement général sur la protection des données à caractère personnel – 2016/679 – RGPD
- La directive sur la distribution d'assurances – DDA – 2016/97
- La directive sur les services de paiement – DSP2 – 2015/2366
- Le règlement établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne – 2018/1807

Nous suggérons l'application des dispositions du RGPD portant sur le « Privacy by design » lors de la conception de l'algorithme, ainsi que les principes de minimisation des risques de violation de données dès lors que ces données sont utilisées par les modèles d'IA (par l'anonymisation et la pseudonymisation). Un équilibre est néanmoins à trouver pour ne pas bloquer l'innovation par des dispositifs trop contraignants.

Question 12 : Analyse des biais

Il s'agit d'analyser les biais pouvant exister aux différents niveaux de mise en place des modèles :

- Dans les données sources.
- Dans les algorithmes.
- Dans les résultats produits.

Le groupe de travail considère que l'analyse des biais doit se faire essentiellement sur les données sources dès la conception de l'algorithme.

Le groupe de travail adhère aux propositions des experts de haut niveau sur l'intelligence artificielle de la commission européenne, publiées dans ses lignes directrices en matière d'éthique pour une IA digne de confiance.

Les lignes directrices de la commission européenne prévoient différentes méthodes d'analyse de biais :

- Méthodes techniques :
    - Architectures pour une IA digne de confiance : « *Les exigences d'une IA digne de confiance devraient être « traduites » en procédures et/ou en contraintes imposées aux procédures, qui devraient être ancrées dans l'architecture du système d'IA. Cela pourrait être accompli au moyen d'un ensemble de règles (comportements ou états) que le système devrait toujours suivre, de restrictions relatives aux comportements ou états que le système ne devrait jamais transgresser, et de combinaisons des deux ou de garanties démontrables plus complexes concernant le comportement du système.* »
    - Éthique et état de droit dès la conception (X dès la conception) : « *Les méthodes destinées à garantir les valeurs dès la conception établissent des liens précis et explicites entre les principes abstraits auxquels le système doit adhérer et les décisions spécifiques de mise en œuvre. L'idée selon laquelle la conformité aux normes peut être incorporée dans la conception du système d'IA est essentielle pour cette méthode.* »
    - Méthodes d'explication:. *pour qu'un système soit digne de confiance, nous devons être en mesure de comprendre pourquoi il s'est comporté d'une certaine manière et pourquoi il a fourni une interprétation donnée ».*

  - Essais et validations : « *Les défaillances des concepts et des représentations utilisés par le système ne sont susceptibles de se manifester que lorsqu'un programme est appliqué à des données suffisamment réalistes.* »
  - Qualité des indicateurs de service : « *Un niveau de qualité approprié des indicateurs de service peut être défini pour les systèmes d'IA, afin de faire en sorte qu'un point de comparaison existe pour déterminer s'ils ont été testés et mis au point en tenant compte de la sécurité et de la sûreté.* »
- Méthodes non techniques
  - Réglementation
  - Codes de conduite
  - Normalisation
  - Certification
  - La responsabilité au moyen de cadres de gouvernance
  - Éducation et sensibilisation pour encourager un état d'esprit éthique
  - Diversité et équipes de conception inclusives
    « il est essentiel que les équipes qui conçoivent, mettent au point, testent, entretiennent, déploient et ou achètent ces systèmes reflètent la diversité des utilisateurs et de la société en général »

<u>Questions 13 et 14 : Rôle de l'IA -Méthodologie et conception</u>

Le problème réside dans l'intégration des algorithmes de l'IA dans le processus existant.

Concernant l'évaluation des résultats, il est nécessaire de mettre en place des processus de contrôle et de back-testing sur tous les logiciels, y compris ceux intégrant des algorithmes développés dans le cadre de projets IA.

Concernant les modèles d'IA, les décisions positives et négatives doivent faire l'objet de tests de cohérence.

L'intégration de l'IA dans le processus métier doit se faire dès la conception (by design).

La différence entre un logiciel classique et un modèle de Machine Learning (IA) réside dans le fait que le risque de dérive est plus important dans le cas d'un modèle de Machine Learning puisque celui-ci est évolutif par nature**. De ce fait, il paraît nécessaire de comparer les résultats de chaque méthode (traditionnelle et IA) et d'analyser les écarts.**

Par ailleurs, il semble nécessaire de back tester les nouveaux modèles lorsque le modèle évolue de manière importante (évolutions de rupture).

L'enjeu sera donc de définir à quel moment le modèle d'IA subit une évolution de rupture.

<u>Questions 15 à 18 : Gestion des risques et Contrôle interne</u>

Les risques liés à l'intégration des modèles de Machine Learning doivent être évalués à l'aune des impacts de ces modèles sur les processus métier. Les risques concernant les algorithmes générant des décisions impactant la clientèle notamment (scoring, distribution de produits financiers…) sont à évaluer et contrôle de manière prioritaire.

Question 19 : Externalisation – Sécurité

Concernant l'externalisation des modèles IA, il nous semble nécessaire d'appliquer les textes règlementaires déjà existants en matière d'externalisation pour la banque, le secteur de l'asset management et de l'assurance:

- Arrêté du 3 novembre 2014 relatif au contrôle interne des entreprises du secteur de la banque, des services de paiement et des services d'investissement soumises au contrôle de l'Autorité de contrôle prudentiel et de résolution ;
- Les orientations de l'EBA sur l'externalisation (EBA/GL/2019/02)
- Les dispositions du Règlement général de l'AMF – RGAMF applicables aux société de gestion de portefeuille
- Instruction n° 2019-I-06 relative à l'information préalable de l'ACPR en cas d'externalisation d'activités ou de fonctions importantes ou critiques et d'évolution importante les concernant

Aspects de sécurité : Adaptation du niveau de sécurité à l'exposition de l'algorithme au monde extérieur. Le groupe de travail considère que les modèles IA sont assez éloignés des portes d'entrée des hackers.

En revanche, les analyses de sécurité doivent porter principalement sur l'injection de fausses données qui auraient un impact défavorable sur la fiabilité des modèles.

Questions 21 à 23 : Approche multifactorielle de l'évaluation

Nous notons la difficulté d'expliquer les modèles par une évaluation analytique. Nous considérons que l'analyse de la cohérence est un bon moyen d'évaluation et est moins complexe à mettre en en œuvre que des modèles concurrents pour lesquels il faudra passer du temps à expliquer les écarts de résultats.

En cas d'utilisation du modèle

- Pouvoir justifier les écarts par rapport au modèle de référence avec la mise en place d'un taux d'acceptation
- Mettre en place un référentiel d'évaluation au niveau des autorités compétentes adapté à chaque catégorie ou modèle (par activité, par taille, par type de données…)
- Expliquer les risques liés aux divergences des modèles avec celui de référence

A chaque création de modèle, il faudrait que les directions de contrôle interne développent en interne des compétentes en IA

Sur la question des modèles concurrents, il nous semble plus efficace d'analyser la cohérence avec les résultats attendus plutôt que l'analyse systématique des écarts de modèles concurrents qui peuvent nuire à la performance.

Savoir conserver des compétences fonctionnelles est particulièrement important (si l'on débranche le pilote automatique, il faut pouvoir piloter l'avion).