Compte-rendu du cycle de conférence " Les lundis de l'IA et de la finance " co-organisés par l'ACPR et Télécom Paris

Astrid Bertrand, Winston Maxwell

▶ **To cite this version:**

Astrid Bertrand, Winston Maxwell. Compte-rendu du cycle de conférence " Les lundis de l'IA et de la finance " co-organisés par l'ACPR et Télécom Paris. 2022. hal-03530291

HAL Id: hal-03530291
https://hal.science/hal-03530291

Submitted on 17 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research paper by

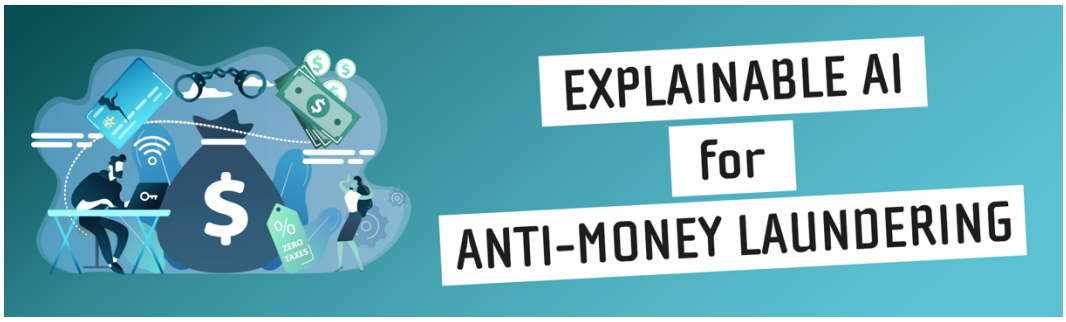
OPERATIONAL AI ETHICS

telecom-paris.fr/en/ai-ethics

Compte-rendu du cycle de conférence « Les lundis de l'IA et de la finance » co-organisés par l'ACPR et Télécom Paris.

Astrid Bertrand[1], Winston Maxwell[2]

Ce cycle de conférences qui s'est tenu de novembre 2020 à mai 2021, a été organisé par les chaires XAI for Anti-Money Laundering et Finance Digitale avec la participation de la chaire DSAIDIS, au sein de l'initiative Operational AI Ethics de Télécom Paris.



Version révisée le 17 janvier 2022

[1] PhD candidate, Télécom Paris, Institut Polytechnique de Paris
[2] Director of Law and Digital Technology Studies, Télécom Paris, i3, Institut Polytechnique de Paris

Co-organisés par l'Autorité de Contrôle Prudentiel et de Résolution (l'ACPR) et Télécom Paris, les "Lundis de l'IA et de la Finance" forment un cycle de conférences autour de la réglementation de l'IA dans le secteur financier. Dans ce cadre, régulateurs, chercheurs et autres acteurs de l'industrie financière se sont réunis un lundi tous les deux mois, de novembre 2020 à mai 2021, pour échanger autour de différentes thématiques du domaine.

En tout, quatre webinaires ont été organisés sur les thématiques suivantes :

- En novembre 2020 : L'explicabilité de l'IA en finance
- En janvier 2021 : L'équité dans les algorithmes
- En mars 2021 : Le partage de données pour l'IA en finance
- En mai 2021 : Réglementation de l'IA dans le secteur financier

1 - L'explicabilité de l'IA en finance

Pour ce tout premier Lundi de l'IA organisé par l'ACPR et Télécom Paris, nous avons abordé le thème de l'IA explicable pour la finance.

David Bounie, professeur d'économie et Olivier Fliche, directeur du pôle Fintech-Innovation à l'ACPR ont introduit le webinaire en présentant le cycle de conférence à venir : les « Lundis de l'IA » sont des conférences se tenant une fois tous les deux mois, sur des thèmes liés à la régulation de l'IA, aux obstacles et aux opportunités du déploiement de l'IA dans la finance, en liant perspectives académiques et perspectives des acteurs du marché.

Le webinaire se déroulait en deux tables rondes. La première présentait le concept d'explicabilité d'un point de vue théorique, en passant par sa définition technique, son utilité pour la société et son application recommandée du point de vue d'un régulateur – l'ACPR. La deuxième table ronde se penchait sur les attentes et les défis du marché concernant l'explicabilité.

Table ronde : l'explicabilité, de la théorie à la régulation

Astrid Bertrand, doctorante à Télécom Paris sur l'explicabilité de l'IA, a introduit cette première table ronde en accueillant **Florence d'Alché-Buc**, professeur en sciences des données à Télécom Paris. Pr. d'Alché-Buc a commencé par nous présenter rapidement l'état de l'art de l'IA. Nous avons été témoins ces dernières années d'avancées significatives dans le domaine de la statistique avec l'apprentissage automatique et symbolique. Les réseaux neuronaux profonds, notamment, ont montré des résultats spectaculaires en matière de performance si bien que l'IA s'intègre dans certains de nos processus décisionnels aujourd'hui. L'explicabilité est la science de décrire l'objectif et le processus de prise de décision de l'outil d'IA d'une manière qui peut être comprise par une personne moyenne. C'est d'autant plus important que les algorithmes d'apprentissage automatique les plus performants sont compliqués à comprendre, de par leurs objectifs de nature statistique notamment. Florence est ensuite revenue sur le besoin d'explicabilité, que l'on peut classifier en 4 catégories : expliquer pour donner confiance, pour contrôler, pour améliorer, et pour découvrir. Selon l'objectif que l'on veut poursuivre, mais aussi selon l'audience de l'explication – expert en data science ou utilisateur profane - les explications peuvent être très variées. Certaines, dites « locales », se contentent d'expliquer une décision algorithmique précise, tandis que d'autres dites « globales » tentent d'expliquer le fonctionnement général du modèle. Les explications dépendent également de la nature des données. Pr. d'Alché-Buc nous a présenté certaines des principales méthodes d'explicabilité utilisées saujourd'hui : LIME, les cartes de saillance, Tree Explainer (SHAP) et les méthodes d'explicabilité dites 'by design'. Pour finir, elle **a souligné que le champ de recherche**

de « l'XAI » est encore à ses débuts, et que les explications aujourd'hui ressemblent plus à des interprétations qu'à des explications. Il faut se demander comment lier les explications avec les raisonnements humains et les structures de connaissance humaines ? Pour le Pr. d'Alché-Buc, rendre un modèle prédictif explicable relève de l'IA symbolique et fait appel à l'automatisation du raisonnement et à la représentation des connaissances.

Télécharger la présentation

Nous avons ensuite accueilli **Winston Maxwell**, directeur d'études en droit numérique à Télécom Paris et responsable du programme Operational AI Ethics. M. Maxwell a commencé par souligner l'importance de l'explicabilité pour construire une IA de confiance, en présentant ce principe comme l'un des 5 pilliers de recherche de l'initiative Operational AI Ethics de Télécom Paris, avec l'équité, l'intérêt général, la responsabilité et la gouvernance. S'appuyant sur un schéma extrait d'un papier de recherche de l'équipe de Télécom Paris[1], M. Maxwell a résumé les facteurs contextuels à prendre en compte pour motiver et orienter le choix d'une explication : l'audience de l'explication, les risques de l'IA, les contraintes légales spécifiques au secteur d'application ou plus générales comme les droits fondamentaux, et enfin les contraintes opérationnelles.
Passant maintenant à la perspective juridique, M. Maxwell a donné un aperçu du droit à l'explication, qui en est encore à ses balbutiements. D'abord, on peut noter une différence entre secteur publique et secteur privé. Dans le secteur public, le droit institutionnel et les droits fondamentaux sont très proéminents, tandis que dans le secteur privé, les lois spécifiques telles que la protection des données personnelles sont plus courantes. Ensuite, M. Maxwell a développé cinq leçons principales du droit à l'explication, s'appuyant sur des cas de jurisprudence et de régulation :

- La première leçon, issue des lois du secteurs public, comme la loi sur la reconnaissance faciale de l'état de Washington, aux États-Unis, ainsi que le règlement européen 'Plateform-to-Business', est le **besoin d'une explicabilité adaptée à la personne qui la reçoit.**
- Le deuxième point s'appuie sur la jurisprudence américaine Houston Teachers, selon laquelle **une décision algorithmique doit être réplicable.**
- Troisièmement, l'affaire Loomis, de l'État de Washington aux États-Unis, nous dit qu'**il faut que les algorithmes** (ici des algorithmes d'évaluation du risque de récidive) **soient accompagnés d'une sorte de notice d'utilisation** (comme pour un médicament) pour mettre en garde contre les risques susceptibles d'être contenus dans leurs prédictions.
- Quatrièmement, un cas de jurisprudence européenne, l'affaire SyRI aux Pays-Bas, met l'accent sur l'importance de **l'auditabilité de l'agorithme pour justifier de sa nature non-discriminatoire.**
- Enfin, La Quadrature du Net, jurisprudence européenne, souligne la nécessité d'une **intervention humaine de qualité pour valider ou infirmer les conclusions des algorithmes**, même quand il est question de mesures contre une menace importante pour la société.

Télécharger la présentation

Le dernier intervenant de cette table ronde « théorique » sur l'explicabilité était **Laurent Dupont**, Data Scientist senior au pôle Fintech-Innovation à l'ACPR. M. Dupont nous a présenté le point de vue du superviseur sur l'explicabilité de l'IA dans le secteur financier, en reprenant des concepts du récent document de réflexion publié par l'ACPR en juin 2020 sur la gouvernance des algorithmes. Le document dégage quatre grands principes de conception de l'IA en finance : la performance, la stabilité, le traitement des données et l'explicabilité. Les objectifs de l'ACPR, en tant que régulateur, ne sont pas de prescrire les outils qui permettent de construire de l'IA mais plutôt de pointer les principes dont l'impact est le plus important.

[1] V. Beaudouin, I. Bloch, D. Bounie, S.Clemençon, F. d'Alché-Buc, J. Eagan, W. Maxwell, P. Mozharovskyi, J. Parekh, 'Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach'.

Une des préconisations du document de réflexion est que l'évaluation d'un algorithme doit couvrir l'ensemble des étapes de son cycle de vie, de sa conception et apprentissage jusqu'à sa validation, son déploiement et sa maintenance, via un processus de contrôle adéquat spécifique à chaque étape. Le document développe également une échelle de 4 niveaux d'explicabilité : du niveau le plus simple, « l'observation » jusqu'au niveau très détaillé de la « réplication ». Le niveau approprié pour un contexte donné se détermine selon les risques de l'application d'IA et l'audience des explications.
M. Dupont a ensuite adopté un angle plus pratique en présentant un tableau des principales méthodologies d'explicabilité qui sont réellement utilisées dans le secteur financier. La première approche, la plus fréquente en finance, est l'explication post-modélisation (SHAP, LIME…) Le deuxième type d'approche consiste à utiliser des architectures explicables qui introduisent dans les algorithmes des règles métier plus logiques à interpréter ou bien y distillent des connaissances « humaines ». L'approche préconisée dans le document de réflexion de l'ACPR est de ne pas déployer de boîte noire si un modèle aussi performant mais plus interprétable existe. Néanmoins, si la boite noire est inévitable, il convient d'être le plus transparent possible sur le gain associé en performance.
Enfin, Laurent Dupont a exposé certaines des limites de l'explicabilité, notamment en citant une étude montrant que les raisonnements humain et algorithmique sont très complémentaires, suggérant par là même qu'ils sont également profondément différents.[2]

Télécharger la présentation

Table ronde : Comment appréhender l'explicabilité pour les acteurs du marché ?

Pour la deuxième table ronde, animée par Olivier Fliche, nous avons eu le plaisir d'accueillir deux experts du marché financier : Jean Barrère, Associé Data Intelligence chez PwC France et Maghreb et Antoine Lissowski, Directeur Général de CNP Assurances, pour nous présenter respectivement les défis du marché pour l'explicabilité et l'explicabilité dans l'assurance.

La présentation de Jean Barrère nous a donné un panorama des applications d'IA dans la banque et des enjeux d'explicabilité associés. L'IA est bien présente dans le secteur financier, mais son développement est contraint par une certaine tension entre la possibilité d'améliorer certains processus, comme la détection des fraudes par exemple, et le besoin de garanties sur le raisonnement logique de l'algorithme. Tous les systèmes d'IA ne sont pas égaux en termes d'explicabilité. On peut établir une distinction claire entre l'IA « boîte noire », performante, qui repose sur des corrélations et l'IA « boite blanche », explicable, grâce à des liens de causalité clairs.
Dans le développement de l'IA, on peut ensuite distinguer deux catégories d'applications de l'intelligence artificielle : celles à moindre risques et celles à haut-risque. Dans la première catégorie, on trouve notamment des systèmes qui fonctionnent en logique boite noire sans exigences explicatives. C'est par exemple tout ce qui tourne autour du service client : la reconnaissance vocale, le routage des appels en temps réel, le NLP pour lire et générer des contrats, la traduction automatique, certains chatbots ou encore l'automatisation de tâches répétitives pour aller chercher des informations alternatives. A contrario, il y a 4 grands cas d'usage à haut risque dans lesquels aucun système d'IA n'est autonome, que ce soit aux États-Unis, ou en Europe : il y a toujours une intervention humaine. Ces applications sont l'évaluation du risque de crédit, la lutte anti-blanchiment d'argent, la détection de fraude et le MiFID (Markets in Financial Instruments Directive), une directive européenne visant à renforcer la protection des investisseurs vis à vis des institutions financières.
Un des principaux défis à relever pour l'explicabilité sera de définir les paramètres pour distinguer les cas d'usage à haut risque et ceux à bas risques. Il faudra également définir des standards minimums d'explicabilité pour les cas à haut-risque, comme la contrainte de réplicabilité ou des contraintes sur les variables utilisées par exemple. Enfin, l'écosystème de l'explicabilité est avant tout

[2] NIST Study Shows Face Recognition Experts Perform Better With AI as Partner, https://www.nist.gov/news-events/news/2018/05/nist-study-shows-face-recognition-experts-perform-better-ai-partner

collaboratif : ces discussions comme celles de ce webinaire entre régulateurs, académiques, institutions financières et cabinets de conseils sont donc essentielles.

Télécharger la présentation

Olivier Fliche s'est ensuite tourné vers Antoine Lissowski en l'interrogeant sur les outils d'intelligence artificielle qui se développent dans le domaine de l'assurance.
M. Lissowski nous a alors fait part du premier contact de l'assurance avec l'IA, qui a d'abord servi à pister les contrats non réclamés pour cause de décès du bénéficiaire. Ensuite l'IA s'est introduite dans des optiques commerciales pour identifier des clients potentiels. Les outils de reconnaissance vocale et optique ont ensuite suivi, et c'est aujourd'hui sur ces techniques que se repose énormément CNP Assurances pour numériser des messages et des documents et les traiter en flux continu sans passer par des opérateurs physiques. Ces applications-là ne soulèvent pas d'interrogations éthiques.
D'autres sont plus concernées par les problématiques d'explicabilité et de transparence. Ce sont les applications de conseil client et de tarification. Celles-ci impliquent – car la loi le requiert – que les sociétés d'assurance récoltent une grande quantité information sur ses clients, pour pouvoir juger de leur appétence au risque par exemple, afin de leur proposer des prestations qui pourront leur être utiles et qui correspondent à un véritable besoin. Le risque est donc que des techniques d'IA indélicates, reposant sur cette base importante de données personnelles, poussent la segmentation de clients jusqu'à des sélections individuelles. À CNP Assurances, une personne a été désignée spécifiquement pour réfléchir à des façons d'identifier les bais de l'IA et de les éviter.
M. Lissowski distingue deux types d'enjeux d'explicabilité pour les cas d'usage de l'IA dans l'assurance :

- Donner les explications suffisantes au client pour qu'il sache pourquoi on lui propose tel produit : c'est un processus déjà très rationnalisé en assurance.
- Donner des explications quant au niveau de tarification. C'est plus compliqué car c'est très dépendant de chaque entreprise d'assurance et de son type de portefeuille de client. Cet enjeu n'est pas nouveau mais avec l'arrivée de l'IA, il faut en plus justifier auprès du superviseur que les outils ne présentent pas de biais.

Liens pour compléter la lecture :

- Document de réflexion, ACPR, Dupont, Fliche and Yang (2020), Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier, juin 2020
- Beaudouin et al. (2020), Flexible and context-specific AI explainability, arXiv: 2003.07703
- Beaudouin et al. (2020), Identifying the 'right' level of explainability, SSRN 3604924
- A.Bertrand, W.Maxwell, X.Vamparys, Do AI-based anti-money laundering (AML) systems violate European fundamental rights?, *International Data Privacy Law*, ipab010, https://doi.org/10.1093/idpl/ipab010

2 - Regards interdisciplinaires sur les biais algorithmiques dans la finance

L'édition du 11 janvier des « Lundis de l'IA et la Finance » de l'ACPR et de Télécom Paris s'est concentrée sur les biais algorithmiques. Le webinaire était la deuxième conférence d'une série dédiée à l'IA dans la finance qui s'inscrit dans les travaux de chaire Digital Finance et Explainable AI for Anti-money Laundering (XAI4AML), comme l'a rappellé David Bounie, professeur d'économie à Télécom Paris. Le sujet choisi pour cette conférence résonne à la fois avec les enjeux d'actualité et avec les intérêts de recherche de Télécom Paris. Notamment, David Bounie est co-auteur, avec Stéphan Clémençon et d'autres chercheurs, d'une étude sur les biais algorithmiques.

Olivier Fliche, Directeur du pôle Fintech-innovation de l'ACPR a ensuite souligné l'importance du sujet des biais pour un régulateur. Les biais statistiques peuvent avoir un impact sur la gestion des risques par les établissements financiers, mais aussi sur la protection de la clientèle, domaines où l'ACPR reste très vigilante.
Le simple fait de parler des biais soulève déjà des défis en matière de vocabulaire. Le terme anglais « fairness » signifie à la fois la loyauté et l'équité, bien que ces termes aient des sens différents en français. Quant au terme « biais », il couvre différents phénomènes, allant des biais de sélection à des biais cognitifs.
Loin de pouvoir couvrir en détail tous les enjeux relatifs aux biais, ce webinaire se veut un échange entre plusieurs acteurs, académiciens et industriels, sur la façon de gérer les biais introduits par les nouvelles technologies. L'étude des biais statistiques est ancienne, mais quels changements induisent les algorithmes ?

Table ronde 1 : Biais algorithmique, équité et discrimination

Pour commencer la première table ronde, animée par David Bounie, Jean-David Fermanian, professeur de Finance et Statistique à l'ENSAE, a présenté le problème du point de vue mathématique. Il souligne la différence entre les concepts statistiques et les approches éthiques des biais. Le défi est de s'assurer qu'un attribut protégé « S » n'affecte pas la probabilité d'une prédiction « Y ». On peut distinguer plusieurs définitions mathématiques de l'équité : la parité démographique, l'« equalized odds » , l'« equal opportunity » et le « lack of disparate treatment », chaque approche reflétant un angle de vue différent par rapport au problème. Pour corriger les biais, il ne suffit pas hélas d'enlever l'attribut protégé « S », car les autres attributs du vecteur « X » peuvent conduire indirectement à l'attribut protégé. Le défi statistique est d'apporter des corrections aux biais sans détruire le pouvoir prédictif du modèle. Les mesures correctives sont de trois types : « pre-processing », qui consiste à apporter des modifications aux données d'apprentissage, en éliminant certaines données, en modifiant les étiquettes, ou les pondérations ; modifier le modèle lui-même pendant la phase de calibration ; enfin « post-processing », qui consiste à modifier les sorties du modèle pour les rendre plus équitables.

Professeur de mathématiques appliquées à Télécom Paris, Stéphan Clémençon a ensuite souligné les difficultés liées aux algorithmes de reconnaissance faciale. Les biais de sélection seront quasi systématiques dans les données utilisées pour l'apprentissage, à savoir les photos de visages, ce qui peut conduire à des différences de performance importantes entre différents groupes de la population. Le niveau de performance est représenté par le ratio entre faux positifs et faux négatifs, ce ratio figurant sur une courbe dite « ROC ». Dans l'idéal, la courbe ROC devrait être identique quel que soit le groupe de la population, mais en réalité cette courbe peut être très différente selon la couleur de peau ou le genre. La non-représentativité des données d'apprentissage n'est pas la seule cause des différences de performance pour les algorithmes de reconnaissance faciale. Même en présence d'un échantillon d'images représentatif, l'algorithme peut avoir du mal à reconnaître certains types de visage, les visages de bébés par exemple. La correction des biais a presque toujours un impact sur le pouvoir prédictif de l'algorithme, nécessitant des solutions de compromis afin de ne pas détruire complètement la performance.

Célia Zolynski, professeur de droit à l'Université de Paris 1 Panthéon Sorbonne, a ensuite apporté un point de vue juridique sur les biais pour clore cette première table ronde « académique ». Elle rappelle que l'article 1 de la Constitution française assure l'égalité devant la loi de tous les citoyens sans distinction d'origine, de race ou de religion, ce qui signifie le même traitement pour tous. Un traitement différencié doit être justifié par une situation différente. Le principe d'égalité et de non-discrimination figure également à l'Article14 de la Convention européenne des droits de l'homme et à l'Article 21 de la Charte des droits fondamentaux de l'UE. Le principe est ensuite développé dans de nombreux textes nationaux, tel que le Code du Travail et le Code Pénal. En droit, on distingue la discrimination directe, un traitement manifestement différencié en fonction du genre ou d'un autre attribut protégé par la loi, et la discrimination indirecte, un traitement en apparence neutre mais qui conduit en fait à une différence préjudiciable pour un certain groupe de la population. Les sources de

discrimination peuvent être endogènes, par exemple un biais de sélection, ou exogènes, à savoir résulter du contexte ou de l'usage de l'algorithme, le biais de l'automatisation par exemple. Pour identifier et remédier aux biais, les études d'impact sont préconisées à la fois par les autorités européennes (recommandation du 8 avril 2020 du Conseil de l'Europe) et par les autorités canadiennes (Algorithmic Impact Assessment).

Table ronde 2 : Le secteur financier face aux biais algorithmiques

La deuxième table ronde de ce webinaire, animée par Laurent Dupont, data scientist senior à l'ACPR, a regroupé différents acteurs du secteur financier.
Florence Picard, Membre du Haut Conseil de l'Institut des actuaires, était la première à présenter l'approche à l'équité dans son métier, en évoquant le problème de la différenciation tarifaire en assurances. Les outils prédictifs peuvent conduire à une individualisation du risque et des tarifs d'assurance, mais cela vient au détriment de la mutualisation du risque inhérente au métier de l'assurance. Il faut trouver un équilibre, un optimum d'équité. La solidarité conduit automatiquement à l'inéquité pour certains groupes. Par exemple les jeunes en bonne santé vont payer plus cher en assurance santé que ce que justifie un calcul de leurs risques, mais cela permet aux anciens, dont la santé est plus fragile, de payer un tarif abordable. Mais quid des comportements dits non-responsables, par exemple des personnes qui fument ? Quelles sont les limites de la solidarité ? L'équilibre tarifaire relève de la responsabilité de chaque assureur, avec certaines limites imposées par la loi, par exemple l'interdiction d'une différenciation tarifaire en fonction du genre.

Vivien Tran-Thien, Director of AI Consulting au sein de Dataiku, a ensuite donné la vision d'une entreprise du numérique. D'abord, il constate que beaucoup d'entreprises étaient focalisées sur le RGPD et commencent seulement maintenant à prendre conscience des risques liés aux biais algorithmiques. La première réaction des entreprises est souvent d'éliminer les données dites sensibles, ce qui est cohérente avec l'approche RGPD. Hélas, cela ne suffit pas, les biais ayant tendance à réapparaître sous une autre forme. Un autre sujet souvent méconnu concerne l'interaction homme/machine et les biais d'automatisation. Comment préserver la qualité de l'intervention humaine face à une suggestion algorithmique ? Un élément important est de faire en sorte que l'algorithme affiche son degré de confiance dans la prédiction. La présentation des variables les plus importantes dans la recommandation est également utile. Enfin, la traçabilité de chaque recommandation est primordiale afin d'analyser les erreurs.

Gaëtan Goldberg, Chargé de mission Numérique, Droits et Libertés au sein du Défenseur des droits, a ensuite pris la parole en soulignant l'importance de la transparence pour lutter contre les discriminations algorithmiques. Il faut en effet que les personnes ciblées par une décision algorithmiques aient conscience qu'elles font l'objet d'un tel traitement afin d'être en mesure de se plaindre d'une éventuelle discrimination. Cette transparence est déjà prévue pour les algorithmes déployés par l'Etat, mais ne s'impose pas avec autant de force aux algorithmes déployés par le secteur privé. Dans le secteur financier, le Défenseur des droits a eu l'occasion de se prononcer sur un refus de crédit à la consommation en raison de l'âge de l'emprunteur : « si le critère de l'âge peut être une donnée pertinente pour le calcul du risque financier, son utilisation dans les règles métier ne saurait être considérée comme adéquate et proportionnée à la finalité du traitement automatisé lorsqu'elle conduit, non plus à calculer un risque, mais à exclure systématiquement de la procédure d'octroi des crédits les clients âgés de plus de 75 ans ».
Le rapport du Défenseur des droits et de la CNIL sur les algorithmes et les discriminations fournit une série de recommandations, dont la poursuite de recherches sur l'identification et sur les mécanismes de correction des biais. Comme l'a souligné la professeure Zolynski, les études d'impact permettent de prendre en considération les risques de discrimination avant la phase d'exploitation.

Enfin, My Chi To, Executive Deputy Superintendent of Insurance au sein du New York State Department of Financial Services, a apporté un regard transatlantique à cette table ronde. Elle explique que le droit de l'Etat de New York interdit d'une part l'utilisation par les assureurs de critères protégés tels que l'origine ethnique, et d'autre part toute tarification qui serait indûment

discriminatoire (unfairly discriminatory). Le NYDFS a expliqué sa politique en matière d'utilisation d'algorithmes dans l'assurance dans sa lettre circulaire n° 1 du 28 janvier 2019. Cette lettre impose aux assureurs de vérifier que les données utilisées par les algorithmes ne conduisent pas à des tarifs indûment discriminatoires, et que ces données ne contiennent pas des critères protégés. Par ailleurs, la loi impose aux assureurs de donner des raisons d'une décision de refus d'assurer un risque. Il faut qu'il existe un lien logique ou intuitif entre le critère utilisé pour la tarification et le risque. Par exemple, le NYDFS interdit l'utilisation du niveau d'éducation comme critère dans l'assurance automobile, sauf si l'assureur peut démontrer la relation logique entre ce critère et le risque (ce qu'aucun assureur n'a essayé de faire pour l'instant). De nombreux groupes de protection des consommateurs souhaiteraient interdire l'utilisation des scores de crédit (credit scoring) comme critère dans la tarification. Le NYDFS étudie actuellement la question.

3 - Le partage de données pour l'IA en finance

Cette 3ème session des lundis de l'IA et de finance avec l'ACPR et Telecom Paris a abordé les questions de partage de données dans l'industrie financière. David Bounie, professeur d'économie à Telecom Paris, a introduit le séminaire en soulignant la dépendance croissante des entreprises aux données pour se constituer un pouvoir de marché. Le partage des données est donc essentiel pour éviter les pratiques anticoncurrentielles. Ce séminaire a examiné les opportunités et les contraintes du partage des données ainsi que les approches techniques permettant de rendre les collaborations possibles.

1ère table ronde : Partager les données pour l'intérêt général

Olivier Fliche, directeur du département fintech-innovation de l'ACPR, a ouvert la première table ronde de ce webinaire portant sur les questions d'intérêt général à partager des données.

Comme premier intervenant, nous avons eu le plaisir d'écouter Kristen Alma, analyste politique au Groupe d'action financière. Le GAFI est l'organisme mondial de surveillance de la lutte contre le blanchiment d'argent et le financement du terrorisme qui fixe les normes internationales en matière de lutte contre le blanchiment d'argent et le financement du terrorisme. Mme Alma a d'abord présenté les opportunités que représente la mise en commun des données pour l'AML-CFT. En combinant des données provenant de différentes sources et parties, les institutions financières peuvent s'appuyer sur un ensemble de données plus complet et augmenter la performance de leurs modèles dans la détection des schémas de blanchiment d'argent, lesquels impliquent généralement plusieurs institutions financières dans plusieurs pays. Le GAFI a publié un rapport détaillé sur le partage des données par les entités privées en 2017. Mme Alma a ensuite présenté les premiers résultats de l'enquête menée par le GAFI sur les technologies déployées par les banques pour le partage des données. Ce rapport, qui doit être rendu public d'ici juin 2021, met en évidence la nature récente des considérations de partage de données pour les institutions, et le conflit qui se produit entre les besoins de partage et les exigences de confidentialité. Il énumère également les technologies envisagées pour surmonter cette barrière (chiffrement homomorphique, preuves à zéro connaissance, calcul multipartite sécurisé...). Le GAFI prévoit une deuxième phase d'étude pour examiner les moyens et les environnements permettant de favoriser le partage des données avec des contraintes de confidentialité.

Ensuite, nous avons accueilli Bertrand Pailhès, directeur de l'innovation et de la technologie à la CNIL, l'autorité française de protection des données. M. Pailhès a présenté les nouvelles

tendances du XXIe siècle en matière de partage des données, notamment la transparence et l'utilisation d'énormes volumes de données. Il a donné un aperçu de la manière dont les autorités de protection des données abordent ces nouvelles questions et développent des initiatives de partage. Il a notamment présenté quatre scénarios de partage qui utilisent des données produites en dehors du cadre du service public pour remplir des missions d'intérêt général : "données privées ouvertes obligatoires", "données privées générales améliorées", "plateforme de réutilisation des données" et "portabilité citoyenne". Il a également souligné que les entreprises sont généralement réticentes à partager leurs données, et que les approches sectorielles sont plus efficaces pour favoriser les collaborations.

Le dernier intervenant à se joindre à cette table ronde était Antoine Dubus, économiste à ECARES à l'Université libre de Bruxelles. Il nous a expliqué les principes économiques du partage des données. Il a commencé par présenter les données comme un atout concurrentiel, en citant Cremer, de Montjoye, Schweitzer (2019) : "La compétitivité des entreprises dépendra de plus en plus de la possibilité d'accéder en temps voulu à des données pertinentes et de la capacité d'utiliser ces données pour développer des applications et des produits nouveaux et innovants." Les entreprises en position dominante sur les données ont donc un avantage indéniable, mais certaines vont même plus loin et créent des pratiques d'exclusion en empêchant l'accès aux données aux autres concurrents (ex. Facebook en 2017). Par conséquent, il ne reste que quelques entreprises dominantes pour façonner la concurrence sur le marché. Dans ce contexte, la mise en commun et le partage des données peuvent contribuer à renforcer la concurrence sur le marché et à stimuler la coopération pour poursuivre des objectifs d'intérêt public général tels que la lutte contre la fraude. Afin de se conformer en même temps à la réglementation sur la protection des données, Antoine Dubus explique qu'il faudra davantage de coordination entre les agences de protection des données et les autorités de la concurrence.

2ème table ronde : Les techniques de partage de données sensibles

La deuxième phase de ce webinaire avait pour but de présenter quelques approches techniques qui rendent possible le partage des données tout en préservant les contraintes de vie privée. Laurent Dupont, senior data scientist à l'ACPR, a mené les débats de cette session. Il a présenté notre première intervenante, le Dr Catuscia Palamidessi, chercheuse au laboratoire LIX-Inria de l'Ecole Polytechnique, qui a déroulé les concepts sous-jacents aux méthodes de confidentialité différentielle.

M. Palamidessi a tout d'abord exposé le compromis qui s'opère entre la vie privée et la notion d'utilité. Afin de développer des modèles précis, qui peuvent être utiles à la société, les algorithmes ont besoin de données pour s'entraîner, parfois au détriment de la vie privée. Il nous faut donc des mécanismes qui optimisent ce compromis entre utilité et vie privée. L'un d'eux est la confidentialité différentielle (DP). Considérons un ensemble de données qui comprend les données de Bob et d'Alice. Le mécanisme ε-DP introduit un peu de bruit statique ε de sorte à être toujours en mesure d'exploiter les résultats des données (les données sont toujours utiles), mais de ne plus être en mesure de différencier les informations de Bob et d'Alice. La confidentialité différentielle standard (alias modèle central) est l'architecture typique du mécanisme DP : on regroupe d'abord tous les ensembles de données en un seul, puis on applique un mécanisme ε-différencié. La confidentialité différentielle locale (LDP) est une autre architecture utilisée par Apple et Google par exemple, dans laquelle chaque entrée de données individuelle n est rendue privée avec un niveau de

confidentialité εn, et l'ensemble de données collecté est le regroupement de ces entrées bruitées. Catuscia Palamidessi a également expliqué les mécanismes k-Randomized Response (kRR) et d-privacy (d représente la distance) et a finalement présenté son approche hybride qui présente un meilleur compromis vie privée-utilité que le LDP.

L'intervenant suivant de cette table ronde technique était Sandrine Murcia, CEO et cofondatrice de Cosmian, un éditeur de logiciels d'entreprise qui fournit un traitement collaboratif confidentiel des données sécurisé par la cryptographie. Elle a expliqué l'approche de Cosmian : fournir aux entreprises un outil qui incorpore des techniques de cryptographie avancées issues de la recherche, afin de tirer parti des résultats de la mise en commun des données avec d'autres parties. Plus précisément, Cosmian propose deux façons de faire de l'informatique collaborative. La première est celle où les entreprises veulent lancer l'informatique collaborative sur des données protégées sans révéler les données propriétaires sous-jacentes. La technique utilisée par Cosmian dans ce cas est le traitement des données entièrement cryptées ("Fully Encrypted Data Processing"). La seconde est destinée aux entreprises qui souhaitent déplacer leurs données vers un autre environnement, comme le nuage, pour y effectuer des calculs, mais qui ne veulent pas que leurs données ou leurs demandes soient partagées avec cet environnement. Le calcul multipartite (MPC) serait alors utilisé.

Laurent Dupont a ensuite accueilli Maxime Agostini, CEO et cofondateur de Sarus, une entreprise dont l'objectif est d'aider les entreprises à extraire de la valeur de toute donnée sensible en utilisant la confidentialité différentielle. Maxime a commencé par exposer le raisonnement qui sous-tend l'utilisation de la confidentialité différentielle. La première idée qui vient à l'esprit pour garantir l'anonymat des données est de simplement supprimer les PII (Personally Identifiable Information). Par exemple, on supprimerait les noms, les dates de naissance et les adresses des clients dans une base de données Netflix. Cependant, en faisant cela, on resterait avec les parties uniques des données, dans notre exemple, la liste ordonnée unique des films regardés par les clients, ce qui permet de retrouver l'identité des clients. Cela a conduit les scientifiques à chercher une meilleure définition, mathématique, de l'information anonyme : la confidentialité différentielle. Cette définition garantit que l'information anonyme ne peut pas mener à des informations sur un individu donné, peu importe ce que l'attaquant sait déjà. Enfin, Maxime nous a partagé l'idée qu'ils aiment mettre en avant à Sarus : oubliez le partage des données, même anonymes, le plus important est de partager la connaissance.

Tous les dossiers de présentation des panélistes seront disponibles ici dans quelques jours, et le replay de l'ensemble du webinaire, y compris les tables rondes, sera également affiché.

Des liens utiles pour compléter la lecture :
https://www.fatf-gafi.org/media/fatf/documents/German-Presidency-Priorities.pdf
https://www.fatf-gafi.org/fr/publications/recommandationsgafi/documents/guidance-information-sharing.html?hf=10&b=0&s=desc(fatf_releasedate)
https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-artificial-intelligence-and-data-protection/
www.cosmian.com
www.sarus.tech

4 - Réglementation de l'IA dans le secteur financier : perspectives croisées en Asie et en Europe

Lors de ce quatrième Lundi de l'IA et de la Finance organisé par l'ACPR et Télécom Paris, nous avons porté un regard international sur la régulation de l'intelligence artificielle dans le secteur de la finance.
David Bounie, professeur d'économie à Télécom Paris, et **Olivier Fliche**, directeur du pôle Fintech-Innovation à l'ACPR, ont ouvert la conférence en présentant les deux croisements de perspectives adoptés dans la conférence : celui des régulateurs et celui des universitaires d'une part et celui de l'Union européenne et celui de l'Asie.
Nous avons eu le plaisir de recevoir deux experts de la réglementation et de la supervision financière, Jan Ceyssens et Xuchun Li, respectivement de la Commission européenne et de l'Autorité monétaire de Singapour, et deux universitaires, John Armour et Douglas Arner, respectivement de l'Université d'Oxford et de l'Université de Hong Kong.

Comme premier intervenant, nous avons accueilli **Jan Ceyssens**, chef de l'unité Finance numérique à la Commission européenne, qui nous a aidés à décortiquer la récente proposition de la Commission européenne sur la réglementation de l'IA. M. Ceyssens a commencé par mettre en lumière le contexte de cette proposition. Dans le cadre de sa stratégie visant à faire de l'Europe un leader sur les questions numériques, la Commission a publié en 2021 un paquet sur la stratégie européenne en matière d'IA avec plusieurs objectifs. Parmi ces multiples stratégies, la proposition définit un cadre juridique pour l'IA, basé sur l'hypothèse générale que l'IA est bénéfique pour l'intérêt public, mais qu'elle comporte des risques qui doivent être gérés. Le cadre juridique propose une approche fondée sur le risque, en distinguant entre les systèmes d'IA présentant des risques inacceptables, les systèmes d'IA à haut risque, les IA soumises à des exigences de transparence spécifiques et les IA à risque minimal ou nul.
Les systèmes d'IA à haut risque doivent respecter un certain nombre d'exigences basées sur deux facteurs : leur impact possible sur la sécurité et sur les droits fondamentaux.
M. Ceyssens a ensuite présenté comment cette proposition s'articulerait dans le secteur financier. La proposition contient un cas d'utilisation dans la finance sur l'évaluation du risque de crédit. Mais la réglementation de l'IA dans la finance va au-delà de ce cadre de surveillance transsectoriel : elle doit être intégrée dans le système de surveillance existant par les autorités de contrôle du secteur financier, notamment l'EBA, l'EIOPA et l'ESA. Les prochaines étapes consisteront en la négociation de la proposition législative par le Parlement européen. Une fois adoptée, il y aura 2 ans de période transitoire avant que le règlement ne devienne directement applicable.

Le deuxième intervenant était Xuchun Li, chef du bureau de développement de l'IA à l'Autorité monétaire de Singapour (MAS). Il a d'abord présenté la charte FEAT du MAS pour le développement de l'IA et ses 4 principes : Équité, Éthique, Responsabilité et Transparence. Ensuite, Dr. Li a décrit l'approche Veritas, qui consiste en une mise en œuvre pratique des principes FEAT. L'objectif est de créer une méthodologie standardisée et modulaire pour la mise en œuvre des principes ci-dessus, dont le code source est rendu public. La phase 1 de l'approche Veritas a eu lieu en 2020 et s'est concentrée sur les principes d'équité et deux cas d'utilisation : l'estimation du risque de crédit et le marketing client. Les résultats sont deux livres blancs ainsi que le code de programmation associé. La phase 2 aura lieu cette année, couvrira les 4 principes FEAT, et élargira les cas d'utilisation à l'assurance et à la détection

des fraudes. Le résultat final de l'approche Veritas sera un livre blanc consolidé présentant des méthodologies pour évaluer l'alignement d'un système avec les principes FEAT.

Nous avons ensuite accueilli **John Armour**, professeur de droit et de finance à l'Université d'Oxford. Sa présentation a porté sur la relation entre la réglementation des produits et la gouvernance d'entreprise, en utilisant comme exemple la proposition de la Commission européenne sur la réglementation de l'IA. Il a expliqué que lorsque de nouvelles technologies sont introduites sur le marché, de nouveaux risques sont également engendrés. Certains sont anticipés par une réglementation appropriée, mais d'autres ne sont pas prévus à l'avance : ce sont les risques émergents. La question est de savoir comment gérer ces risques émergents. Une approche est le principe de précaution, largement utilisé dans le domaine pharmaceutique, qui vise à restreindre le déploiement de la nouvelle technologie jusqu'à ce qu'il soit prouvé qu'elle peut être utilisée en toute sécurité. Mais il faut être proportionné dans l'application de ce principe, afin de ne pas renoncer aux avantages que l'IA peut apporter.
La proposition de l'UE s'efforce de faire face aux risques émergents : certains cas d'utilisation de l'IA sont interdits, des normes minimales sont fixées pour les applications à haut risque, et la proposition prévoit également un système permettant de faire passer un système d'IA dans la catégorie à haut risque si de nouveaux problèmes apparaissent. Ce régime encourage également les fournisseurs à élaborer volontairement des normes pour les systèmes classés à haut risque et à autoréguler les systèmes à faible risque. Cette approche semble s'inspirer des notions déjà existantes de réglementation des produits et de « gouvernance produit ». D'une part, la réglementation des produits doit être calibrée selon des analyses coûts/bénéfices qui tiennent compte de l'évolution rapide de l'innovation dans le domaine. D'autre part, la « gouvernance produit » impose aux entreprises de démontrer l'intérêt de leur produit pour le consommateur. Pour une bonne gouvernance interne des produits, il faudra garantir une approche axée sur les avantages pour le consommateur et éviter de se contenter de cocher les cases d'exigences minimales.

Le dernier intervenant de la conférence était **Douglas Arner,** professeur de droit à l'université de Hong Kong. Il a commencé par souligner que la finance est l'un des secteurs réglementés les plus numérisés et qu'à ce titre, elle représente un terrain très propice au développement de l'IA. Le secteur concentre des volumes massifs de données, de communications et de puissance de calcul et s'appuie déjà sur des moyens analytiques sophistiqués. En outre, un large éventail de spécifications réglementaires spécifiques est déjà en place dans la finance, ce qui crée un contexte de règles et d'objectifs précis où l'IA peut exceller. En outre, les acteurs financiers sont les plus grands dépensiers en matière de technologie, poussés par une course constante à la meilleure IA, comme dans le cas du trading. Cependant, l'IA introduit de nouveaux risques dans la finance : pour la stabilité financière, pour la cybersécurité ou pour l'innovation car l'IA fonctionne mieux dans une concentration massive de données qui favorise les situations "winner takes all". Plusieurs méthodologies de régulation sont possibles, notamment l'autorisation ou l'identité électronique... Le Dr Arner a souligné une approche fondée sur la responsabilité humaine ("accountability"). Pour conclure, il a mis en évidence certaines différences entre juridictions sur les approches aux données. La Chine, qui reposait à l'origine sur une approche aux données fondée sur la propriété, se tourne aujourd'hui progressivement vers une approche de type "pool de données", dans laquelle de grandes masses de données seraient mises à la disposition d'un plus grand nombre d'acteurs pour soutenir le développement de l'IA. Dans l'UE, le marché restreint fortement l'utilisation des données afin de protéger les droits fondamentaux.

La deuxième partie du webinaire était consacrée à une table ronde modérée par David Bounie et Olivier Fliche. Les éléments de la discussion comprenaient :

- les différences d'approche entre l'Asie et l'Europe
- comment lier la réglementation transversale et sectorielle : il est toujours nécessaire d'adopter des mesures plus larges et globales, par exemple pour protéger les droits fondamentaux, qui doivent ensuite être intégrées dans les structures existantes.
- l'utilisation des cadres de responsabilité personnelle, qui existent déjà dans les secteurs réglementés, pour accroître la responsabilisation,
- l'importance des interactions entre les entreprises réglementées et les régulateurs (expérimentations, notamment)
- la façon dont nous évaluons les performances d'une IA : est-ce par rapport à un objectif, par rapport au modèle le plus performant ou par rapport aux performances humaines ?
- les avantages et les risques de l'approche "human-in-the-loop".
-

Tous les dossiers de présentation des panélistes seront disponibles ici dans quelques jours, avec le replay du webinaire.

Liens utiles pour compléter la lecture :

- Papiers sur la réglementation de l'IA écrits par les intervenants :
  - '**Artificial Intelligence in Finance: Putting the Human in the Loop**', Dirk A. Zetzsche, Douglas W. Arner, Ross P. Buckley, Brian Tang (2020)
  - '**FinTech, RegTech and the Reconceptualization of Financial Regulation**', Douglas W. Arner, Janos Nathan Barberis, Ross P. Buckley (2016)
  - '**Taking Compliance Seriously**', John Armour, Jeffrey N. Gordon, Geeyoung Min *(2020)*
  - '**The RegTech Book:** the financial technology handbook for investors, entrepreneurs and visionaries in regulation'
- Plus de contenu sur le sujet :
  - '**Consumer-Lending Discrimination in the FinTech Era**', Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace
  - '**Introduction to FinTech**': The world's most popular FinTech online course, with edX