AI Act et règlementation sectorielle : interactions et conséquences sur la supervision

Julien URI (ACPR)

Françoise Guebs (Banque Nationale de Belgique)

National Bank of Belgium
Eurosystem

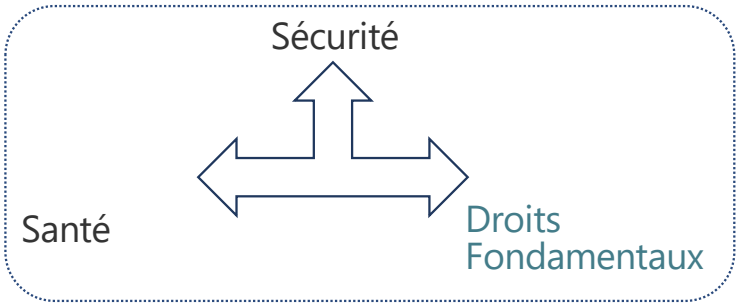
ACPR
BANQUE DE FRANCE

Introduction et structure de la présentation

1. La protection des droits fondamentaux dans la réglementation financière
2. L'articulation entre AI Act et réglementation financière
    - Les exigences applicables selon l'AI Act
    - Focus : l'architecture de supervision => un superviseur unique ?
    - Le cas des modèles internes
    - Les frictions potentielles
3. Quelles perspectives en matière de supervision ?

AI Act et protection des droits fondamentaux

Intérêts Publics Protégés:



Droits Fondamentaux dans l'AI Act

- Droit à la dignité humaine
- **Respect de la vie privée et familiale**
- **Protection des données à caractère personnel**
- Liberté d'expression et d'information
- **Le droit à la non-discrimination**
- **La protection des consommateurs**
- Le droit des travailleurs
- Les droits des personnes handicapées
- **L'égalité de genre**,
- Les droits de propriété intellectuelle
- **Le droit à une bonne administration**
- **Le droit à un niveau élevé de protection de l'environnement etc. (Considérant n°48 et Charte des Droits fondamentaux de l'UE)**

- L'AI Act poursuit un double objectif: (1) promouvoir l'investissement, la recherche et l'innovation, la compétitivité et la croissance, et (2) éthique de l'IA, centrée sur l'humain (en ce compris la protection des droits fondamentaux):
  - **La gravité de l'atteinte aux droits fondamentaux (en considération de la gravité du préjudice et de la probabilité de sa survenance sous-tend la classification d'un système d'IA comme "à haut risque"** (Article 7)
  - **Approche de la législation "sécurité des produits" (New Legislative Framework),** en ce compris (1) une opérationalisation via la **normalisation**, et (2) une **approche de gestion des risques** très différente de l'approche de protection des droits fondamentaux de la Charte : les risques à identifier et gérer ne concernent que ceux qui peuvent être *raisonnablement atténués ou éliminés dans le cadre du développement ou de la conception du système d'IA à haut risque*, **ou** *par la fourniture d'informations techniques appropriées*; notion d'acceptation d'un risque résiduel global... (Article 9)
- L'AI Act est sans préjudice du droit de l'Union en vigueur, en particulier en ce qui concerne la protection des données, la protection des consommateurs, les droits fondamentaux, l'emploi et la protection des travailleurs, et la sécurité des produits, qu'il vient **compléter (Considérant n°9)**.

Réglementation prudentielle et protection des droits fondamentaux (avant l'AI Act)

- Le superviseur prudentiel poursuit traditionnellement d'autres **objectifs** que la protection des droits fondamentaux, en particulier la protection de la stabilité financière … même si **la protection des droits fondamentaux reste pertinente sur le plan prudentiel (risques réputationnels, risques juridiques)**.

- Comme les Etats membres, **le superviseur prudentiel est *lié par* le droit de l'Union protégeant les droits fondamentaux** (Traité de l'Union Européenne, Traité de Fonctionnement de l'Union Européenne, Charte des Droits Fondamentaux de l'Union européenne)

- Cette mission est toutefois différente de celle des autorités de protection des droits fondamentaux à désigner selon l'article 77 de l'AI Act qui ont un ***mandat spécifique pour superviser ou faire respecter*** les obligations du droit de l'Union visant à protéger les droits fondamentaux.

- Même en l'absence de disposition spécifique, **le superviseur prudentiel ne peut ainsi ignorer le droit de l'Union protégeant les droits fondamentaux. La réglementation prudentielle** peut d'ailleurs imposer des exigences congruentes (par ex: qu'un raisonnement économique sous-tende l'utilisation de certaines variables).

Réglementation prudentielle et protection des droits fondamentaux (avant l'AI Act) (exemples)

Orientations de l'ABE sur l'octroi et le suivi des prêts

"Lorsqu'ils utilisent des innovations fondées sur les technologies à des fins d'octroi de crédit, les établissements doivent::

[…] comprendre la qualité des données et des entrées du modèle, ainsi que ***détecter et prévenir les biais dans le processus de décision en matière de crédit***, en veillant à ce que des garanties adéquates soient mises en place pour assurer la confidentialité, l'intégrité et la disponibilité des informations et des systèmes".

EBA report on Big Data and advanced analytics & Follow-up report

"The importance of data protection in the context of BD&AA needs also to be reflected appropriately at the organisational and management levels of institutions. In particular, ***institutions need to comply with the GDPR throughout the entire lifecycle of a BD&AA application (e.g. the development and production processes) when using personal data for training models or for other purposes during the steps in the BD&AA process****".*

"Another important aspect of a trustworthy model is its fairness. ***Fairness requires that the model ensures the protection of groups against (direct or indirect) discrimination.*** *Discrimination can affect in particular smaller populations and vulnerable groups (e.g. discrimination based on age, disability, gender reassignment, marriage or civil partnership, pregnancy or maternity, race, religion or belief, sex, sexual orientation, etc.).* ***To ensure fairness (non-discrimination), the model should be free from bias".***

"a model running in a production environment can be regularly monitored to ensure that it has not deviated into discriminatory behaviour. In addition, ***it should be noted that having a diverse workforce (i.e. composed of a good balance of men and women and people from different backgrounds and with complementary skills) can also help to ensure the early detection of bias/discrimination issues and represents, therefore, a competitive advantage in building fair solutions****".*

Articulation avec la réglementation sectorielle : principes généraux

- De nombreuses règles et attentes ont déjà vocation à encadrer le développement et les attentes prudentielles relatives à l'usage de l'IA dans le secteur financier, moyennant mise en oeuvre adaptée aux spécificités de l'IA, par exemple:
  - Règles générales relatives à la gouvernance et la gestion des risques
  - Règles spécifiques relatives aux modèles internes et guide BCE sur les modèles internes
  - DORA
  - Orientations de l'Autorite Bancaire Européenne (ABE) du 29 novembre 2019 sur la gestion des risques liés aux TIC et à la sécurité (EBA/GL/2019/04) ...
- Principe d'économie conceptuelle et législative qui prévaut en droit de l'Union (nonobstant la perception de "sur régulation", *mais* spécificité de l'IA en tant que technologie d'usage général)
- L'AI Act tient compte de ces deux dimensions dans ses principes d'articulation de la législation sectorielle et des nouvelles obligations qu'il impose

Articulation avec la réglementation sectorielle : principes d'articulation dans l'AI Act

- Les exigences de l'AIA (sous réserve de leur opérationalisation via les standards harmonisés) et celles applicables aux institutions financières semblent **en partie congruentes**.
- Pour le secteur financier, l'AI Act prévoit, afin de réduire la complexité et les couts et de favoriser la cohérence, **l'intégration de certaines exigences procédurales dans les exigences sectorielles**, ou que **la conformité à la réglementation sectorielle (*moyennant prise en considération des standards harmonisés*) "vaut respect des obligations" de l'AI Act**.
- En outre, l'AI Act reconnait **la complémentarité** de ses règles et d'autres règles de l'Union. Ces règles pourraient alors être substituées à celle de l'AI Act si elles fournissent un niveau de protection équivalent, ou être appliquées de façon combinée.

(1) Intégration dans la réglementation financière

Fournisseur:

- Système de gestion de la qualité (art. 17(4), avec des exceptions limitées)
- Documentation (art. 18(3))
- Tenue des *logs* générés automatiquement (art. 19(2))

Déployeur:

- Obligation de surveillance (art. 26(5))
- Tenue des *logs* générés automatiquement (art. 26(6))

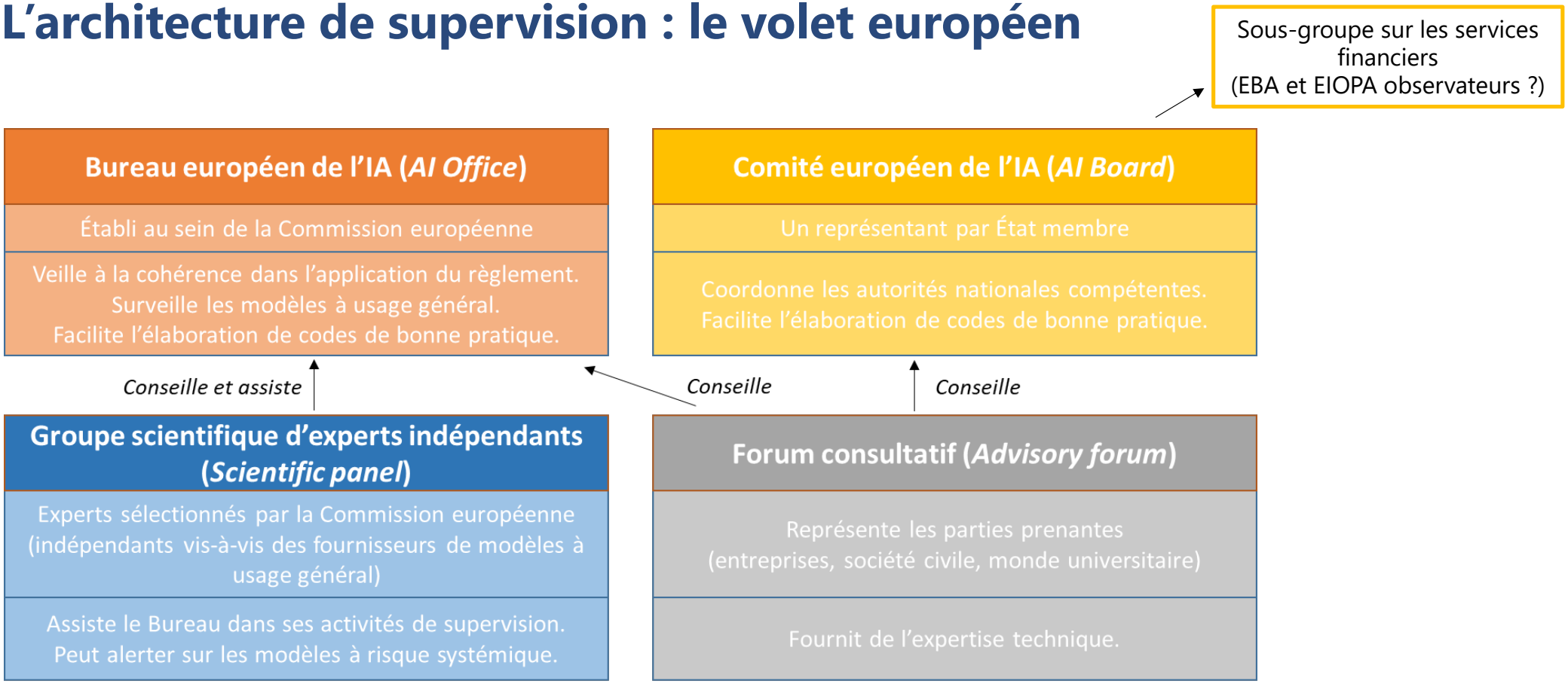
(2) Intégration ou complémentarité

Fournisseur:

- Intégration ou combinaison avec les procédures de gestion des risques (art. 9(10))
- Signalement d'incidents graves (art.73(9))
- Surveillance après commercialisation (opt-in) (art. 72(4))

Règles et orientations sectorielles: CRD/CRR, DORA, etc.

L'architecture de supervision : le volet européen



L'architecture de supervision : le volet national

- L'essentiel du contrôle des obligations de l'AI Act **revient aux États membres**
- Ils désignent à cet effet des **autorités de surveillance du marché** : en principe les **superviseurs financiers**, même si les États conservent la liberté de s'organiser comme ils l'entendent
- **Périmètre de surveillance :**
  - Systèmes d'IA :
    - Systèmes à haut risque en lien direct avec la fourniture de services financiers
    - Systèmes à risque limité ?
  - Acteurs :
    - Acteurs financiers, y compris les établissements de crédit de taille significative (SI) => remet en question, pour le secteur bancaire, le principe du contrôle par le superviseur prudentiel
    - Mais possiblement aussi des fournisseurs non financiers, que les autorités de surveillance du marché pourraient soumettre à leur contrôle si nécessaire

Autorités de surveillance du marché : principales missions

- **Rôle** : assurer la conformité des produits au moyen d'une démarche fondée sur les risques, en effectuant des contrôles appropriés et d'une ampleur suffisante (règlement européen « conformité des produits » 2019/1020)
- **5 missions principales**, en combinant AI Act et RCP :

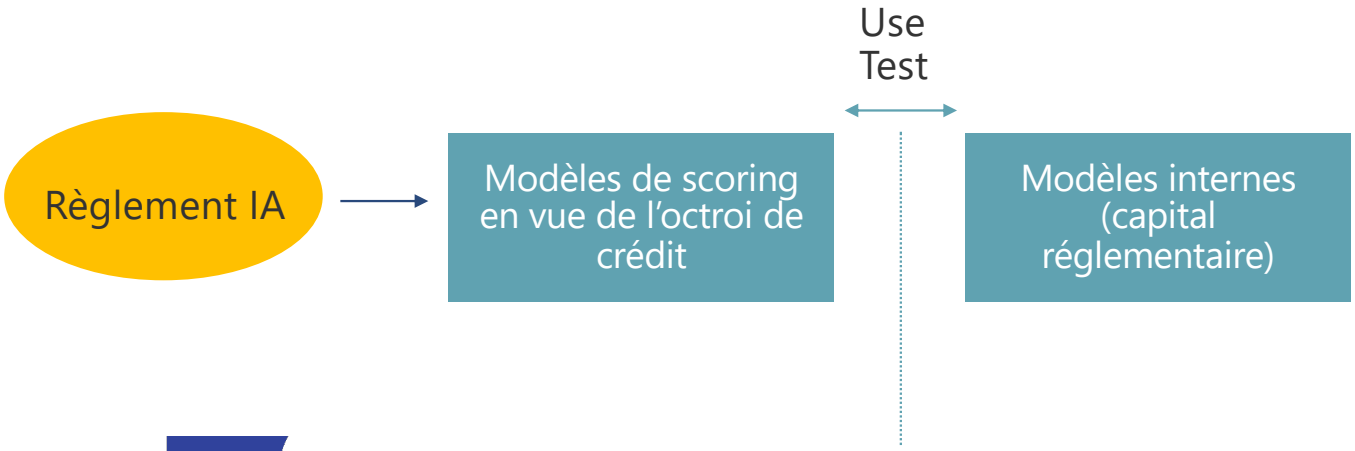
Collecte d'informations et surveillance du marché	Évaluation de « premier niveau » de la conformité	Contrôles approfondis de conformité	Gestion de la non-conformité et sanctions	Coordination
• Collecter les analyses d'impact sur les droits fondamentaux • Traiter les rapports d'incident et les réclamations	• Des systèmes d'IA (et des différents opérateurs) • Sur la base de la documentation disponible mais sans besoin d'accéder au code informatique	• Impliquant notamment l'accès au code du système	• Prescription de mesures correctives aux opérateurs • Prise directe de mesures correctrices • Instruction des dossiers de sanction	• Avec les autorités de la concurrence ; • Avec la BCE ; • Avec les autres MSA en France et en Europe ; • Avec l'*AI Office* (IA à usage général)

Autorités de surveillance du marché : pouvoirs

- **Requérir des opérateurs les informations nécessaires** à l'évaluation de conformité d'un système d'IA...
  - ... y compris les **jeux de données** utilisés pour développer le système (Art. 74.12 de l'AI Act)
  - ... et le **code source** (Art. 74.13), si cela est approprié et nécessaire
- Ouvrir des **enquêtes**
- Procéder à des **inspections inopinées sur place**
- Exiger que les opérateurs prennent les **mesures appropriées** pour mettre fin à une non-conformité… voire, en cas de non-exécution, **prendre directement lesdites mesures** (y compris, le cas échéant, l'interdiction d'un système)
- Imposer des **sanctions**

Articulation avec la réglementation sectorielle: le cas particulier des modèles internes IRB

- Les systèmes d'IA utilisés à des fins prudentielles pour le calcul des besoins en fonds propres sont exclus du périmètre du Règlement IA (Considérant (58)) mais... les exigences de l'AI Act vont nécessairement avoir un impact sur ces modèles internes du fait de l'exigence du "Use Test" (art. 144 (1) (b) CRR) en vertu duquel les notations internes et estimations de défaut utilisées dans le calcul des exigences de fonds propres doivent jouer un role essentiel dans le processus décisionnel/l'approbation des crédits



Articulation avec la réglementation sectorielle: frictions potentielles? (Equité algorithmique)

- Une tension pourrait survenir des attentes en matière d'équité algorithmique
- Mais cela va dépendre de l'approche retenue par les standards harmonisés, qui est très débattue...
- Exemple typique de la perpétuation des schémas historiques de discrimination induits par la différence de rémunération homme/femme (évoqué dans le considérant (58) de l'AI Act) et l'utilisation de la rémunération comme variable – faut-il promouvoir une approche transformante des biais pour favoriser un accès égal au crédit? Si oui, laquelle? Est-ce le rôle des banques et des superviseurs ou des Etats Membres et Cours et Tribunaux? (voir notamment: *S. Wachter, B. Mittelstadt and C. Russell: Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law)*
- Toutefois, une approche de différenciation des emprunteurs en fonction du risque de crédit – et notamment de la rémunération – n'est pas seulement protectrice de la stabilité financière et de la solvabilité des banques, mais également des emprunteurs (prévention du surendettement)

Articulation avec la réglementation sectorielle: frictions potentielles? (Transparence / explicabilité)

- L'AI Act semble déployer une approche holistique de l'explicabilité en concentrant ses exigences sur la transparence du système d'IA, la documentation technique (explicitant notamment la 'logique générale' qui guide le processus de décision du système) facilitant la surveillance humaine, intégré dans un système de gestion des risques *(voir C. Panigutti, Hamon, Hupon et al., The role of explainable AI in the context of the AI Act)*

- Il pourrait ne pas exiger la transparence "by design" (IA nativement interprétable) ou l'explicabilité du comportement du modèle via certaines techniques (IA Explicable, "XAI") (Voir notamment articles 13/14, Annexe IV):

  - La conception et le développement des systèmes d'IA à haut risque sont tels que **le fonctionnement de ces systèmes est suffisamment transparent *pour permettre aux déployeurs d'interpréter les sorties d'un système* et de *les utiliser de manière appropriée***

  - *Le système doit être fourni au déployeur en telle manière que les personnes physiques chargées d'effectuer un contrôle humain ont la possibilité: de **comprendre correctement les capacités et les limites pertinentes du système d'IA à haut risque** et d'être en mesure de surveiller correctement son fonctionnement, y compris en vue de détecter et de traiter les anomalies, les dysfonctionnements et les performances inattendues; d'avoir conscience d'une éventuelle tendance à se fier automatiquement ou excessivement aux sorties produites par un système d'IA à haut risque (**biais d'automatisation**), en particulier pour les systèmes d'IA à haut risque utilisés pour fournir des informations ou des recommandations concernant les décisions à prendre par des personnes physiques; **d'interpréter correctement les sorties du système d'IA à haut risque, compte tenu par exemple des outils et méthodes d'interprétation disponibles***

  - *Exemple de mesures soutenant l'explicabilité: **précision du contexte et des conditions spécifiques d'utilisation via la notice d'utilisation et la documentation technique**, précision du **type de formation nécessaire pour l'humain qui assure la surveillance du système (quelles sont ses caractéristiques, limitations etc.) et pour prévenir les biais d'automatisation***

- Les attentes prudentielles pourraient aller plus loin: *voir notamment ACPR, Modèles internes des banques pour le calcul du capital réglementaire (IRB) et intelligence artificielle (Mars 2024) qui débat de la conformité de modèles boite noire couplée à des modèles d'interprétabilité ex post et l'intérêt de modèles nativement interprétables au regard des exigences d'interprétabilité déduites de la CRR; futur guide BCE sur les attentes relatives aux modèles internes IA pour l'IRB*

Articulation avec la réglementation sectorielle : prochaines étapes

- Le superviseur comme les établissements auront besoin que le **cadre réglementaire soit entièrement clarifié** :
  - Nécessité de réaliser une **cartographie complète** des exigences de l'AI Act pour les institutions financières, et de leurs interactions avec le reste de la réglementation
    - ➢ **Ce travail a commencé au sein des ESAs**, en coordination avec la Commission européenne et la BCE
    - ➢ Il fera aussi l'objet d'un sous-groupe du Comité européen de l'IA (*AI Board*) en 2025
  - Une difficulté : le contenu de certaines exigences dépend de **normes harmonisées** que les **organismes européens de normalisation** (CEN-CENELEC) doivent publier à échéance (probable) fin 2025

- **Action du superviseur** : l'AI Act ne précise pas l'articulation avec le reste de la réglementation du secteur financier :
  - Vise à laisser la liberté de s'organiser au mieux
  - Toutefois, des orientations des ESAs pourraient à terme s'avérer utiles pour préciser les choses

Vers la supervision (1/2)
Maîtrise des risques liés à l'IA : pas de révolution copernicienne

- **Les établissements financiers apparaissent bien outillés pour relever le défi de la conformité à l'AI Act :**
  - Culture de la maîtrise des risques, et dispositifs de **gouvernance** et de **contrôle interne** qui peuvent être répliqués ou étendus au champ de l'IA (moyennant parfois quelques adaptations, ex : suivi des modèles dans le temps)
  - Nombre d'exigences de l'AI Act sont **déjà couvertes** par d'autres dispositions règlementaires (exemple : risque cyber / DORA)
- **C'est bien dans cet esprit que l'ACPR entend contrôler les dispositifs de maîtrise des risques liés à l'IA** :
  - Pour la conformité à l'AI Act :
    - Surveillance de marché **ex post**, et approche **basée sur les risques**
    - Utilisant au maximum les **synergies avec le contrôle prudentiel**
  - Vaudra également « hors » de l'AI Act => systèmes d'IA non classés à « haut risque » mais jugés critiques par le superviseur

Vers la supervision (2/2)
Ne pas sous-estimer, cependant, le défi posé par certains aspects de l'IA

- Exemples les plus évidents : **explicabilité, équité**
- Pour les établissements :
  - **Se doter des capacités humaines et techniques** pour pouvoir apporter la preuve que les différentes exigences réglementaires sont respectées
  - L'ACPR entend s'assurer que **la maîtrise des risques est effective** : éviter le *box ticking*, et vérifier au contraire que les algorithmes sont gérés et surveillés par des personnes compétentes qui en comprennent le fonctionnement profond
- Pour le superviseur :
  - Se doter d'une **doctrine** sur certains sujets « nouveaux »
  - Développer une **méthodologie** ad-hoc de **l'audit de l'IA (**par exemple autour des 5 dimensions suivantes : performance et robustesse, explicabilité, équité, vie privée et confidentialité des données, cyber-sécurité)
  - Construire des **synergies** avec les autres superviseurs de l'IA (en France et en Europe)

Conclusion

- L'IA constitue le premier moteur de transformation du secteur financier aujourd'hui ; les enjeux sont donc importants
- Pour autant, aux yeux de l'ACPR et de la BNB, les principes de saine gestion des risques et de gouvernance demeurent inchangés ; ils constituent notre boussole.
- L'IA - et l'AI Act - comportent cependant des aspects nouveaux, qui vont demander une montée en compétence, pour les établissements comme pour le superviseur
- L'ACPR et la BNB partagent ainsi de nombreux défis avec les établissements financiers : il convient donc d'avancer ensemble pour les surmonter