ÉTUDE DE DOSSIER

Cadre de direction 2026

L'intelligence artificielle connaît un développement rapide qui transforme l'ensemble de l'économie, des processus industriels aux services financiers. Les banques centrales ne sont pas épargnées par cette évolution : elles explorent déjà l'utilisation de l'IA pour améliorer leurs prévisions économiques, renforcer leurs capacités d'analyse ou encore accroître l'efficacité de leurs opérations. Dans le secteur financier, les établissements bancaires et les investisseurs recourent également de plus en plus à des modèles d'intelligence artificielle pour la gestion des risques, l'allocation des actifs ou la détection des fraudes. Toutefois, ces avancées soulèvent également de nombreuses interrogations. L'opacité de certains modèles, la concentration des infrastructures technologiques entre les mains de quelques acteurs ou encore les risques de comportements mimétiques pourraient créer de nouvelles vulnérabilités pour le système financier. Dès lors, l'intelligence artificielle doit-elle être considérée avant tout comme une opportunité pour renforcer l'efficacité des banques centrales et du secteur financier, ou comme une nouvelle source de risque systémique nécessitant une adaptation des cadres de supervision et de régulation ?

À partir des documents ci-joints, vous répondrez aux questions suivantes :

1. Quelles sont les principales évolutions engendrées par le développement de l'intelligence artificielle dans la sphère économique et financière ?
2. Par quels canaux les modèles d'intelligence artificielle affectent les principales missions des banques centrales et des superviseurs : stabilité des prix, stabilité financière et supervision ?
3. Quels sont les risques émergents liés à l'utilisation de l'intelligence artificielle par le secteur financier ?
4. En quoi le développement de l'intelligence artificielle soulève-t-il, en Europe, des enjeux de souveraineté économique, financière et technologique ? Quels risques spécifiques peuvent résulter du recours à des solutions d'intelligence artificielle non souveraines ?
5. How can central banks and supervisory authorities integrate artificial intelligence into their analytical, monitoring and supervisory toolkit? How do you assess the main opportunities and limitations of these developments? *(Réponse à rédiger en langue anglaise)*

Les questions sont indépendantes ; nous vous recommandons toutefois de les traiter dans l'ordre. Il n'est pas nécessaire de recopier l'intitulé des questions. Les réponses doivent être rédigées en langue française (à l'exception de la question n° 5).
Une attention particulière sera portée à la qualité de la rédaction.

LISTE DES DOCUMENTS JOINTS

1. **Les enjeux de l'IA du point de vue de la Banque Centrale – Discours de Denis BEAU**
   www.banque-france.fr – 02/04/2026 – 4 pages
2. **Les banques centrales doivent se préparer à l'impact profond de l'IA sur l'économie et le système financier : BRI**
   www.bis.org – 25/06/2024 – 2 pages
3. **Face à l'IA offensive, la BCE pousse les banques à revoir leur copie en cybersécurité**
   www.solutions-numeriques.com – 28/05/2026 – 2 pages
4. **L'IA pousse les banques centrales à repenser inflation et taux**
   https://fr.euronews.com – 29/04/2026 – 4 pages
5. **Navigatin uncertain times with the help of artificial intelligence**
   www.ecb.europa.eu – 21/04/2026 – 4 pages
6. **Special topic – Artificial intelligence**
   www.eba.europa.eu – Novembre 2024 – 9 pages
7. **L'intelligence artificielle : bénédiction ou malédiction pour la transformation du secteur financier ?**
   www.banque-france.fr – 8/11/2024 – 4 pages
8. **Deepfakes : l'image de la Banque de France détournée pour piéger les entreprises**
   www.latribune.fr – 28/04/2026 – 2 pages
9. **The Financial Stability Implications of Artificial Intelligence**
   www.fsb.org – 14/11/2024 – 1 page
10. **Financial Stability Risks Mount as Artificial Intelligence Fuels Cyberattacks**
    www.imf.org – 07/05/2026 – 3 pages
11. **Les enjeux de l'intelligence artificielle pour le système financier**
    www.banque-france.fr – Juin 2024 - 14 pages
12. **Financial stability in the age of artificial intelligence: the role of algorithmic architecture**
    www.ecb.europa.eu – 21/05/2026 – 4 pages
13. **Agents of change – Speech by Sarah Breeden**
    www.bankofengland.co.uk – 30/06/2026 – 4 pages
14. **Souveraineté numérique : De l'urgence d'organiser la coopération entre le public, le privé et les communs numériques**
    www.conseil-ia-numerique.fr – Mai 2026 – 2 pages

1

Les enjeux de l'IA du point de vue de la Banque Centrale

Discours de Denis Beau, Premier sous-gouverneur

Brest, 2 avril 2026

Mesdames et Messieurs, chers étudiants,

Je voudrais tout d'abord remercier les organisateurs de l'Institut d'Administration des Entreprises de Brest pour leur invitation à ouvrir cet évènement consacré à l'Intelligence Artificielle et ses enjeux, en vous présentant ses enjeux du point de vue d'une Banque Centrale comme la Banque de France, une institution publique indépendante de près de 9 000 personnes, membre de l'Eurosystème et du Mécanisme de Supervision Unique des banques européennes, qui a pour objectifs d'assurer la stabilité monétaire, la stabilité financière et de fournir des services à l'économie et à la société.

Pour la Banque de France l'IA affecte à la fois les objectifs qu'elle poursuit et la gestion des moyens qu'elle mobilise pour les atteindre et je vous propose donc de vous présenter successivement les enjeux auxquels est confrontée la Banque du fait de ces deux types d'impacts.

1. Pour pouvoir assurer la stabilité des prix, la Banque de France, avec les autres banques centrales de l'Eurosystème, a besoin de comprendre et de mesurer quelle va être l'ampleur de sa diffusion au sein de l'économie et son impact. Nous suivons en effet de près tous les développements susceptibles d'influencer la conjoncture économique des années à venir ainsi que les prix. Concernant **l'impact de l'IA sur le PIB, même si celui-ci est complexe à appréhender car il emprunte une variété de canaux de demande et d'offre et sa quantification est difficile, je retiens des données récentes et les travaux d'études les trois observations suivantes** :

    - **Ses effets dans l'immédiat sur l'investissement sont substantiels**, notamment aux États-Unis, à l'avant-garde du développement de l'IA. Le FMI observe que depuis début 2024 le dynamisme de l'investissement des économies avancées du G20 s'explique principalement par celui du secteur américain des semiconducteurs et des logiciels. En France, l'investissement dans les logiciels et les bases de données a doublé au cours de la dernière décennie (x 2,5 aux États-Unis), tandis que l'investissement dans la construction de centres de données a été multiplié par 2,5.

    - **À long terme, l'IA devrait avoir un impact significatif sur la productivité.** D'après Cerutti et al (2025), la croissance mondiale devrait être accrue de 0,1 à 0,4 p.p./an, en raison de la hausse de la productivité de 0,1 à 0,2 p.p./an (amplifiée de ses effets sur l'accumulation du capital et sur la demande), avec de larges écarts selon les pays. Ainsi selon Aghion et Bunel (2024), l'adoption de l'IA au cours de la prochaine décennie pourrait augmenter la croissance annuelle de la productivité des pays développés de 0,1 à 1,2 p.p./an selon le rythme et l'ampleur de l'adoption, avec une estimation médiane d'environ 0,7 p.p. Cette fourchette est globalement cohérente avec les estimations de l'OCDE pour la France comprises entre +0,3 et +1,0 p.p./an.

- **Pour l'instant, l'adoption de l'IA a eu des effets limités sur l'emploi total et ses implications pour les inégalités restent incertaines.** D'après Aghion et al. (2025), l'analyse de données françaises portant sur la période 2017-2020 indiquent que les effets de l'adoption de l'IA devraient être positifs à horizon 4 ans. Selon une enquête récente (Yotzov et al., 2026), **les déplacements de travailleurs devraient s'intensifier dans les années à venir, y compris dans les secteurs de services** employant des travailleurs peu qualifiés tels que le commerce de détail et l'hôtellerie.

En agissant sur l'économie par une variété de canaux de demande et d'offre, l'IA rend très incertain son effet global sur l'inflation et sur r*, le taux d'intérêt nominal permettant de maintenir une inflation stable avec un niveau de demande compatible avec le plein emploi. De plus, les effets de l'IA peuvent être sur le niveau mais aussi sur la volatilité de l'inflation.

Quelles conséquences tirer de ces observations pour la conduite de la politique monétaire ? À ce stade aucune ne s'impose pour ce qui concerne la politique monétaire pour la zone euro. L'impact de l'IA constitue l'un des nombreux ajustements structurels en cours (comme le changement climatique ou le vieillissement de la population) qui affectent l'offre de long terme de l'économie et la croissance potentielle. Elle représente un élément important de l'évaluation du côté de l'offre, mais ne présente pas de caractéristiques spécifiques justifiant une modification de la fonction de réaction de la politique monétaire.

L'IA a en revanche d'ores et déjà un impact sur notre mission de stabilité financière, du fait de sa diffusion et de ses modalités d'utilisation par les intermédiaires financiers, et en particulier ceux que l'Autorité de contrôle prudentiel et de résolution (ACPR), le « gendarme » des banques et des assurances en France, qui est adossé à la Banque de France, a la tâche de superviser afin de limiter leur probabilité de défaillance. En effet, comme l'a montré une enquête récente de l'ACPR, la quasi-totalité des banques et des organismes d'assurance en France utilise désormais l'IA, que ce soit pour améliorer et personnaliser les services aux clients, pour optimiser les processus internes ou pour mieux gérer les risques.

Pour autant, l'adoption croissante de l'IA dans le secteur financier présente un certain nombre de **risques**, **en particulier** pour la **stabilité financière -**que l'on pense par exemple à la dépendance des acteurs financiers aux grands fournisseurs de modèles d'IA, qui sont aussi les fournisseurs majeurs de services de cloud- et, de manière évidente, pour les **consommateurs**.

Ces risques justifient **l'encadrement règlementaire** de l'utilisation de l'IA, qui doit permettre un développement maitrisé. Ce cadre, c'est bien évidemment le **règlement européen sur l'IA**, mais **c'est aussi la règlementation sectorielle**, qui s'applique à l'IA comme à n'importe quelle technologie utilisée par les acteurs financiers.

Les superviseurs de l'ACPR sont donc dès maintenant confrontés au défi de mettre sur pied une **surveillance efficace et efficiente des systèmes d'IA**. Celle-ci doit être à la fois sélective -ce que nous appelons l'approche « basée sur les risques »- et approfondie, c'est-à-dire que nous devons avoir la capacité d'« ouvrir le capot » des algorithmes afin d'en examiner les caractéristiques techniques. Pour cela, nous allons devoir **monter en compétence -**y compris en recrutant de nouveaux talents- et nous atteler à développer une **méthodologie d'évaluation de l'IA** dans le secteur financier traitant notamment les nouveaux enjeux comme **l'explicabilité** et **l'équité** des algorithmes. Les superviseurs doivent enfin **accompagner** les établissements du secteur financier pour les aider à développer les « bons » outils de maîtrise des risques.

2. Outre ses objectifs l'IA impacte également la gestion des moyens que la Banque de France met en œuvre pour les atteindre et soulève de ce fait également des enjeux en matière de gestion interne que je vous propose de vous présenter maintenant.

En effet, l'IA n'est plus un horizon lointain pour la Banque de France : **c'est un outil opérationnel, déjà ancré dans notre quotidien.** Pour cela nous avons construit un socle technologique robuste : une plateforme de données pleinement opérationnelle, et des compétences internes solides, allant de la data science à l'ingénierie IA. Ce socle nous a permis de déployer des cas d'usage bien réels, au service de nos missions. En 2025, nous avons aussi déployé Copilot Chat sur l'ensemble des postes : un assistant conversationnel qui a offert à tous nos agents un premier aperçu du potentiel de l'IA dans leurs tâches quotidiennes dans un cadre strict de sécurité, limité aux données non sensibles.

Nous entrons désormais dans une phase d'accélération résolument positive. Les métiers expriment un intérêt croissant pour ces technologies : les attentes sont nombreuses, et les idées affluent de toutes parts pour transformer les pratiques. Avec les expertises déjà présentes en interne, que nous allons encore renforcer, nous disposons d'un environnement idéal pour amplifier l'usage de l'IA et en faire un levier structurant de modernisation. Pour accompagner cet élan, nous avons lancé, en février 2026, la mission « **Banque Innovante** ». Elle poursuit trois objectifs simples : **assurer une gouvernance claire** de l'IA, **simplifier et automatiser** nos processus métier et surtout, **rendre l'IA accessible à tous**, et pas seulement à quelques experts.

Cette mission s'appuie sur une feuille de route structurée en 2 volets. Le 1er volet vise à développer l'IA pour tous, en accompagnant chaque collaborateur vers un usage quotidien et éclairé de ces outils. **Faire adopter l'IA** par tous les collaborateurs constitue un enjeu :

- **d'employabilité**. Un programme de formation riche est déjà en place contribuant à garantir la montée en compétences nos collaborateurs, et ainsi assurer leur employabilité à long terme. Il sera renforcé prochainement par des dispositifs d'accompagnement ciblés, adaptés aux activités et aux domaines métier.
- **de productivité**, afin de gagner en productivité dans toute l'organisation par l'automatisation des tâches redondantes et chronophages, l'augmentation de la concentration des équipes sur les tâches à forte valeur ajoutée et le partage, la réutilisation des actifs IA déployés par les collaborateurs.
- Et **de culture** en vue de libérer la capacité des équipes à s'auto-transformer.

Le second volet vise à identifier et mettre en œuvre les actions de transformation sur nos **priorités structurantes, l'efficience opérationnelle** (temps agent libéré, gains d'ETP, amélioration de la productivité), **la faisabilité technique et la qualité des données** (fraîcheur, disponibilité, potentiel de déploiement à grande échelle, intégration au SI), ainsi que **la valeur ajoutée pour le métier** (amélioration de la qualité, résolution d'irritants métier, facilitation du quotidien des utilisateurs).

Comme vous le voyez l'IA est pour nous désormais un **levier stratégique** pour renforcer notre capacité à remplir nos missions régaliennes et moderniser notre fonctionnement. Nous bâtissons ainsi, très concrètement, la **banque centrale de demain.**

Permettez-moi de vous donner un **exemple très concret** de cette démarche. Pour nos missions de **supervision**, à l'ACPR, nous examinons les données issues du *reporting* des établissements, afin d'identifier les grandes tendances comme les signaux faibles. Il s'agit donc d'un **terrain idéal**

pour déployer des outils -nous les appelons « *SupTech* » dans notre jargon- permettant aux équipes de contrôle de gagner en efficacité.

Au-delà d'améliorer la productivité, l'ambition de l'ACPR est de doter ses agents de **nouvelles capacités** : ne pas remplacer les humains par des machines, mais créer des équipes de **superviseurs augmentés capables de faire à la fois plus et mieux**. Nous avons par exemple élaboré un outil appelé « **Véridic** », un **LLM** capable d'extraire les caractéristiques des produits commercialisés en assurance-vie à partir de leur « document d'information clé » afin de les classer en fonction de leur niveau de **complexité**, et donc de leur **risque** pour les souscripteurs.

Au-delà des outils développés, les expérimentations de l'ACPR permettent de dégager un certain nombre d'**enseignements** pour le déploiement de l'IA : elles montrent par exemple qu'avant de déployer des outils d'automatisation, il convient bien souvent de **simplifier en amont**, tant les concepts que les processus, car la technologie n'a pas vocation à gérer l'ambiguïté que la complexité génère inévitablement. Un autre enseignement, c'est qu'il convient non seulement de **former** les agents à l'utilisation des nouveaux outils -par exemple en leur apprenant à rédiger des prompts efficaces- mais aussi, dans le même temps, de les **sensibiliser aux risques** que ces outils comportent.

Je finirai mon propos en soulignant **les enjeux de souveraineté dans nos choix en matière d'IA.**

Nous sommes actuellement en phase de choix d'une solution pour mettre à disposition de nos agents un **assistant IA capable de traiter des données sensibles**, intégré à nos outils, souverain et sécurisé.

Lorsqu'une banque centrale adopte l'IA, la question centrale n'est pas seulement : « Quelle technologie est la plus performante ? », mais aussi : **« Cette technologie nous permet-elle de rester maîtres de nos données et de nos systèmes ? »**

Nous manipulons des informations sensibles : données bancaires, prudentielles, statistiques économiques. Elles doivent rester protégées en toutes circonstances : personne d'extérieur ne doit pouvoir y accéder, et aucun acteur privé ou public ne doit pouvoir nous en couper l'accès. Or l'IA repose largement sur des briques technologiques non européennes, ce qui crée des risques concrets : dépendance à un seul fournisseur, exposition à des lois extraterritoriales, augmentation unilatérale de tarifs, ou arrêt d'un service essentiel. Pour une banque centrale, c'est un **risque de perte d'autonomie**.

C'est pourquoi la souveraineté est devenue un critère essentiel dans tous nos choix d'IA.

Nous appliquons une règle simple : **plus une donnée est sensible, plus elle doit rester dans nos infrastructures internes ou dans un cloud de confiance européen.** Nous nous méfions du verrouillage technologique : chaque solution doit être **réversible** et substituable. Enfin, la souveraineté, c'est aussi la **résilience** : une IA n'est utile que si l'infrastructure qui la porte reste disponible, sécurisée et contrôlée.

En résumé, pour une banque centrale, l'IA n'est pas seulement un enjeu d'innovation : c'est un enjeu d'**autonomie**, de **maîtrise** et de **continuité de mission**. Nous devons innover, oui, mais jamais au prix de notre souveraineté.

2

Les banques centrales doivent se préparer à l'impact profond de l'IA sur l'économie et le système financier : BRI

- Les banques centrales devraient chercher à exploiter au mieux l'intelligence artificielle (IA), en anticipant son impact sur l'économie et le système financier et en l'exploitant dans leurs propres opérations. Une adoption généralisée pourrait avoir des répercussions sur la dynamique de l'inflation.
- Le secteur financier est l'un des plus exposés aux avantages et aux risques de l'IA. Les avantages comprennent des améliorations en matière de prêts et de paiements ; les risques comprennent des cyberattaques plus sophistiquées.
- L'importance accrue des données en tant que pierre angulaire de la révolution de l'IA accélère la nécessité d'une coopération entre les banques centrales.

L'adoption rapide de l'IA impose aux banques centrales de s'approprier cette nouvelle technologie, a déclaré aujourd'hui la Banque des règlements internationaux (BRI), invitant les responsables politiques à anticiper les effets transformateurs de l'IA sur l'économie et à l'utiliser pour affiner leurs propres outils d'analyse en vue d'assurer la stabilité financière et la stabilité des prix.

Le chapitre spécial du *Rapport économique annuel 2024 de la BRI* expose les implications des nouvelles applications de l'IA pour les banques centrales. L'IA est sur le point d'avoir un impact sur le système financier, les marchés du travail, la productivité et la croissance économique. Avec une adoption généralisée, elle pourrait renforcer la capacité des entreprises à ajuster plus rapidement les prix en réponse aux changements macroéconomiques, avec des répercussions sur la dynamique de l'inflation. Le travail des banques centrales en tant que gardiennes de l'économie sera également directement affecté en tant qu'utilisatrices de première ligne des outils d'IA.

Les cas d'utilisation de l'IA par les banques centrales comprennent l'amélioration des prévisions à très court terme *(nowcasting)* en utilisant des données en temps réel pour mieux prévoir l'inflation et d'autres variables économiques et pour examiner les données afin de détecter les vulnérabilités du système financier, permettant aux autorités de mieux gérer les risques. Les données sont devenues une ressource encore plus précieuse avec l'avènement de l'IA et seront la pierre angulaire de l'utilisation de la technologie par les banques centrales.

Les modèles d'IA de nouvelle génération ont captivé notre imagination collective grâce à leurs surprenantes capacités, mais ils ont également une incidence directe sur la manière dont les banques centrales font leur travail. D'énormes quantités de données pourraient fournir des informations plus rapides et plus riches pour détecter des modèles et des risques latents dans l'économie et le système financier. Tout cela pourrait aider les banques centrales à mieux prévoir et piloter l'économie.

Hyun Song Shin, responsable de la recherche et conseiller économique à la BRI

Les effets sur la demande et donc sur les pressions inflationnistes dépendront de la rapidité avec laquelle les travailleurs déplacés pourront trouver un nouvel emploi et si les ménages et les entreprises anticipent correctement les gains futurs de l'IA. À court terme, l'offre pourrait dépasser la demande, ce qui pourrait réduire les pressions, mais ces effets pourraient s'inverser au fil du temps, la demande se rattrapant elle aussi grâce à l'augmentation des revenus. Les banques centrales devront rester attentives à cette dynamique dans leur politique monétaire.

Dans le secteur financier, l'IA peut améliorer l'efficacité et réduire les coûts des paiements, des prêts, de l'assurance et de la gestion d'actifs, selon le rapport. La BRI a averti que l'IA introduit également des risques, tels que de nouveaux types de cyber-attaques, et peut amplifier les risques existants, tels que les comportements de masse, paniques bancaires, ou des ventes d'urgence.

Le pôle d'innovation de la BRI teste les capacités de l'IA dans plusieurs domaines en collaboration avec des banques centrales partenaires.

Les banques centrales ont été les premières à adopter l'apprentissage automatique et sont donc bien placées pour tirer le meilleur parti de la capacité de l'IA à imposer une structure à de vastes quantités de données non structurées.

Par exemple, le projet Aurora explore comment détecter les activités de blanchiment d'argent à partir des données de paiement et le projet Raven utilise l'IA pour améliorer la cyber-résilience, pour n'en citer que deux de notre portefeuille.

Cecilia Skingsley, responsable du pôle d'innovation de la BRI

Note aux rédacteurs :

- *Le pôle d'innovation de la BRI vise à développer les biens publics dans l'espace technologique.*
- *Le rapport économique annuel 2024 et le rapport annuel 2023/24 de la BRI seront publiés le 30 juin.*

Face à l'IA offensive, la BCE pousse les banques à revoir leur copie en cybersécurité

Les capacités émergentes de certains modèles d'intelligence artificielle inquiètent les superviseurs bancaires. Entre réunions de crise, appels à l'investissement et échanges d'informations avec des établissements américains déjà exposés à ces outils, la Banque centrale européenne tente d'accélérer la préparation du secteur face à des attaques automatisées capables d'exploiter des vulnérabilités à une vitesse inédite.

Selon Reuters, la Banque centrale européenne pousse désormais les banques de la zone euro à renforcer leurs investissements en cybersécurité face à des outils capables d'exploiter automatiquement des vulnérabilités logicielles.

La BCE hausse le ton sur le risque cyber lié à l'IA

La Banque centrale européenne multiplie les échanges avec les banques de la zone euro autour des risques liés aux nouveaux modèles d'intelligence artificielle. Le sujet était au cœur d'une réunion organisée le 26 mai par Frank Elderson, vice-président du Conseil de surveillance de la BCE, qui a rassemblé plus de 300 représentants du secteur bancaire, d'institutions publiques et d'associations professionnelles, selon une source proche du dossier citée par l'AFP.

Depuis plusieurs semaines, le superviseur européen interroge les établissements sur leur niveau de préparation face à l'émergence d'outils capables d'identifier et d'exploiter automatiquement des vulnérabilités logicielles. Pour Luis de Guindos, vice-président sortant de la BCE, le sujet dépasse désormais le cadre de la cybersécurité traditionnelle. Interrogé par Reuters le 27 mai, il estime que les banques devront augmenter leurs investissements afin de renforcer des défenses appelées à devenir "structurelles" dans les prochaines années.

"La cybersécurité devient de plus en plus importante", a-t-il déclaré, en insistant sur la nécessité d'investissements "pervasifs", y compris pour les petits établissements.

Mythos, le modèle qui alimente les inquiétudes des superviseurs

Au centre des préoccupations figure Mythos, un modèle développé par Anthropic. Selon les explications relayées par Frank Elderson dans un entretien publié sur le site de la BCE, ce système serait capable de "découvrir et exploiter de manière autonome des vulnérabilités à une vitesse et à une échelle inédites".

Toujours selon le responsable de la BCE, le modèle pourrait également combiner plusieurs failles mineures pour construire des attaques complexes auparavant réservées à des profils experts. Il serait aussi capable d'analyser des correctifs de sécurité pour en déduire rapidement de nouvelles vulnérabilités exploitables, réduisant le délai d'exploitation de plusieurs semaines à quelques heures.

“L’absence d’accès ne constitue pas une excuse pour l’inaction”, a rappelé Frank Elderson dans les colonnes du Financial Times. Les banques européennes n’ont pas directement accès à Mythos, réservé pour l’instant à un nombre limité d’organisations américaines. Mais certaines filiales européennes de groupes américains supervisés par la BCE, comme JPMorgan Chase, y ont accès.

Une pression croissante autour de la résilience opérationnelle

Pour la BCE, la question dépasse désormais la seule solidité financière des établissements. Face à des cyberattaques jugées plus fréquentes et plus sophistiquées, les superviseurs mettent l’accent sur la capacité opérationnelle des banques à maintenir leurs services en cas d’incident.

Le dialogue engagé par la BCE doit se poursuivre dans les prochains mois afin de favoriser le partage d’informations et la préparation de plans d’action communs. La mobilisation ne se limite d’ailleurs pas à l’Europe. Selon l’AFP, le secrétaire américain au Trésor Scott Bessent et le président sortant de la Réserve fédérale Jerome Powell ont eux aussi réuni plusieurs grandes banques en avril pour évaluer les risques liés au modèle développé par Anthropic.

Charlotte Rabatel

4

L’IA pousse les banques centrales à repenser inflation et taux

Les banques centrales tentent de comprendre comment l’IA redessine l’inflation, sans parvenir à un consensus.

Depuis près de trois ans, les banques centrales ont abordé l’intelligence artificielle comme le changement climatique ou la démographie : une force de long terme qu’il faut suivre de près, mais pas encore un outil de politique monétaire.

Cette distinction n’a désormais plus lieu d’être.

Les responsables de politique monétaire commencent à présenter l’IA comme une mutation structurelle de l’ampleur de l’électrification ou d’internet, qui va remodeler l’inflation, les taux d’intérêt et même les instruments dont disposent les banques centrales pour les fixer.

Le débat ne porte plus sur l’importance de l’IA.

Il porte sur le calendrier, les canaux de transmission et le sens du mouvement : à quelle vitesse ses effets se matérialiseront, si les prix monteront ou baisseront d’abord, et comment les banques centrales doivent réagir face à une force potentiellement inflationniste à court terme mais désinflationniste à plus long terme.

Ce que la BCE et la Bundesbank font déjà avec l’IA

La Banque centrale européenne est celle qui a le plus rapidement fait passer l’IA de la théorie à la pratique.

Dans un **billet de blog** ***(source en anglais)*** publié le 21 avril 2026, quatre économistes de la BCE -Óscar Arce, Karin Klieber, Michele Lenza et Joan Paredes- ont révélé que, depuis fin 2022, un modèle d’apprentissage automatique fait partie de l’arsenal d’analyse utilisé pour préparer les décisions de politique monétaire du Conseil des gouverneurs.

Ce modèle s’appuie sur quelque 60 indicateurs couvrant les anticipations d’inflation, les pressions sur les coûts, l’activité réelle et les conditions financières, et il est actualisé plusieurs fois par trimestre.

Ses résultats ont déjà été testés en temps réel.

Aux deuxième et quatrième trimestres 2025, le modèle a signalé des risques de hausse de l’inflation sous-jacente qui se sont ensuite matérialisés, les chiffres définitifs s’établissant environ 20 points de base au-dessus des projections officielles de l’Eurosystème.

« L’intelligence artificielle (IA) peut aider à suivre en temps réel les risques pesant sur l’inflation », écrivent les auteurs.

La Bundesbank suit une trajectoire similaire.

Lors d'une conférence conjointe Bundesbank–SUERF à Francfort, le 9 décembre 2025, le président de la Bundesbank, Joachim Nagel, a affirmé que la banque centrale allemande utilise déjà un large éventail d'applications d'IA pour affiner ses analyses et soutenir les processus de travail.

Cela va des assistants intelligents basés sur le texte à l'analyse automatisée de documents, en passant par un modèle baptisé MILA qui évalue les communications des banques centrales de la zone euro.

« La technologie doit, en fin de compte, être au service des personnes. Et il en va de même pour nous, banques centrales : nous utilisons l'IA pour remplir notre mandat du mieux possible », a déclaré Nagel.

Ce que la Fed dit de l'IA : de la curiosité au cœur du débat

À la Réserve fédérale, l'évolution est moins opérationnelle que conceptuelle, et elle apparaît de plus en plus urgente.

Les responsables sont passés d'une simple reconnaissance de l'IA à des discussions sur la manière dont elle reconfigure les arbitrages fondamentaux de la politique monétaire.

L'an dernier, le gouverneur de la Fed Christopher Waller a soutenu que l'IA est adoptée plus rapidement que les ordinateurs personnels, internet ou les smartphones, et que la question de la productivité se trouve désormais au cœur du débat sur la politique monétaire.

« Une question cruciale est de savoir si l'IA contribuera à un regain de la croissance de la productivité. Toute progression durable de la productivité au-delà de 2 % tendra à soutenir la hausse des revenus réels et du niveau de vie sans pression inflationniste. En tant que décideur monétaire, j'espère que l'IA tiendra cette promesse », a déclaré Waller.

Lors de l'événement Euro20+ organisé par Nagel en novembre 2025, le vice-président de la Fed, Philip Jefferson, a mis en avant l'effet ambivalent de l'IA sur l'inflation.

D'un côté, cette technologie pourrait réduire les coûts de production grâce aux gains de productivité. De l'autre, elle pourrait renchérir le prix des intrants.

« L'IA pourrait exercer une pression à la hausse sur certaines catégories de prix, à mesure que de nombreuses entreprises chercheront à déployer massivement cette technologie. L'IA nécessite également des centres de données, qui se retrouvent en concurrence avec d'autres processus de production pour l'accès au foncier, à l'énergie et à d'autres intrants. Je pense donc que l'impact de l'IA sur l'inflation ne se résume pas à une force désinflationniste », a estimé Jefferson.

La voix la plus politisée dans ce débat est celle de Kevin Warsh, nommé par Donald Trump pour prendre la tête de la Fed à l'expiration du mandat de Jerome Powell en mai.

Warsh a qualifié l'essor de l'IA de vague la plus favorable à la productivité de sa génération et a comparé la période actuelle à la fin des années 1990, lorsque Alan Greenspan avait maintenu une politique plus accommodante que ne l'indiquaient les règles mécaniques et en avait été récompensé par une productivité en hausse et des prix stables.

Mais son audition de confirmation, la semaine dernière, a révélé un profil plus prudent.

Il a décrit l'IA comme une force proche de la « vitesse de libération » et averti que les responsables ne peuvent pas encore compter sur ces gains de productivité.

« L'IA devient si déterminante qu'elle s'approche de quelque chose comme une vitesse de libération », a déclaré Warsh, en avertissant que la Réserve fédérale pourrait devoir repenser ses modèles.

S'il reconnaît que la vague actuelle d'innovations pourrait, avec le temps, atténuer les pressions sur les prix et faciliter la lutte contre l'inflation, il a mis en garde contre le manque de visibilité des responsables sur la façon dont ces gains se répercuteront sur l'emploi, l'autre pilier du mandat de la Fed.

Wall Street se divise entre tenants de la désinflation et faucons de l'investissement

Tandis que les banques centrales débattent encore de la manière d'interpréter l'IA, Wall Street a déjà commencé à la faire entrer dans ses paris de marché. Les plus grands gestionnaires d'actifs et les économistes des banques d'investissement intègrent cette technologie dans leurs prévisions d'inflation, de croissance et de taux obligataires, et la place new-yorkaise s'est scindée en deux paris opposés.

Les partisans de la thèse désinflationniste voient dans l'IA un choc d'offre positif : des prix plus bas, des taux plus faibles et des actifs risqués en hausse.

Les faucons des dépenses d'investissement y voient au contraire un problème d'inflation à court terme : un cycle d'investissement record qui fait grimper les prix de l'électricité, entame l'équilibre épargne-investissement et pousse à la hausse les rendements de long terme avant même l'arrivée des gains de productivité.

Le scénario désinflationniste le plus offensif vient du secteur de la gestion d'actifs. Mike Hunstad, responsable de la division de gestion d'actifs de 1 400 milliards de dollars de Northern Trust, a déclaré au Financial Times en avril 2026 que l'IA pourrait s'avérer l'un des plus grands chocs d'offre positifs de l'histoire économique récente.

Selon Hunstad, si l'IA se traduit par un gain durable de productivité, elle accomplira le travail désinflationniste que des années de politique restrictive n'ont pas permis d'achever.

« C'est presque comme si l'IA devenait votre politique monétaire, et qu'elle allait être plus efficace que tout ce que la Fed, ou d'ailleurs n'importe quelle banque centrale dans le monde, peut faire », estime Hunstad.

Les faucons de l'investissement défendent la thèse inverse : un cycle d'investissement d'une ampleur suffisante pour faire monter les prix de l'électricité et les rendements avant même que les gains de productivité ne se manifestent.

En février, Ben May et Daniel Harenberg, d'Oxford Economics, ont soutenu qu'abaisser les taux de manière préventive en partant du principe que l'IA sera désinflationniste serait une erreur, l'IA alimentant actuellement l'inflation via les prix de l'électricité, les investissements dans les centres de données et les effets de richesse liés à la hausse des marchés boursiers.

« L'impact de l'IA sur l'inflation dépendra de la mesure dans laquelle le soutien apporté au côté offre de l'économie sera compensé par une éventuelle hausse de la demande globale associée », écrit Oxford Economics dans sa note.

Goldman Sachs aboutit à une conclusion similaire. Les économistes Manuel Abecasis et Hongcen Wei se sont penchés sur ce qui est peut-être le canal de transmission le plus sous-estimé entre l'IA et l'inflation : les prix de l'électricité.

L'inflation de l'électricité aux États-Unis a atteint 6,9 % sur un an jusqu'en décembre 2025, bien au-dessus de l'inflation globale PCE, à 2,9 %. Goldman anticipe une inflation de l'électricité pour les ménages proche de 6 % en 2026 comme en 2027, avant un ralentissement autour de 3,5 % en 2028.

« Nous nous attendons à ce que les centres de données dopent fortement la demande d'électricité, représentant environ 40 % de la hausse totale de la demande d'énergie au cours des cinq prochaines années », explique Abecasis.

Goldman estime que la hausse des coûts de l'électricité ajoutera 0,2 point de pourcentage à l'inflation globale en 2026 et 0,15 point en 2027, les secteurs de la santé, des transports et de la restauration absorbant l'essentiel de la répercussion indirecte sur les prix sous-jacents.

La question porte sur le calendrier, pas sur l'issue

Un consensus large se dégage désormais sur un point : l'IA est d'une ampleur telle qu'elle oblige les banques centrales à repenser le fonctionnement de l'économie.

Ce qui reste à trancher, c'est l'ordre d'apparition des effets. Si les gains de productivité arrivent en premier, les banques centrales pourraient disposer de marges pour baisser les taux sans raviver l'inflation.

Si c'est la vague d'investissements qui se matérialise en premier, via les prix de l'énergie, la demande de capital et la valorisation des actifs, les responsables qui assoupliraient trop tôt pourraient être contraints de faire marche arrière.

Cela marque une rupture nette avec la situation d'il y a quelques années à peine, lorsque l'intelligence artificielle n'apparaissait quasiment pas dans les discours des banques centrales.

Piero Cingari

5

Navigating uncertain times with the help of artificial intelligence

Artificial intelligence (AI) can help track inflation risks in real time. A new ECB model based on machine learning informs experts how likely it is that inflation will be much higher or much lower than they expect.

In times of growing economic and political uncertainty, prices can change more rapidly and more strongly. This is why monetary policy decisions rely not only on the most likely path for inflation, which economists like to call the "baseline", but also on an assessment of the risks surrounding it.[1] In other words, how likely it is that inflation will turn out to be higher or lower than the baseline. Estimating this has become much more complicated in a world of growing uncertainty. This blog post shows how the ECB is applying new AI-based tools to cope with this increased complexity.

The Eurosystem (the ECB and the national central banks of the euro area countries) utilises a comprehensive toolkit to analyse the risks surrounding inflation forecasts.[2] For example, market-based sensitivity analysis focuses on how surprise developments in key economic variables, such as oil prices, would affect the euro area economy. In addition, model-based analysis examines how a wide range of potential developments could shape the entire distribution of future macroeconomic outcomes. This offers a broader perspective than the sensitivity analysis. However, the standard economic models used in this context are typically based on only a handful of economic indicators. They may also rely on restrictive assumptions.[3]

The machine learning model described in this blog post comes with two major advantages for the risk analysis.[4] First, it is able to handle a greater number of economic indicators. Second, machine learning models are able to capture very general and complex data patterns. This allows it to detect and cope with non-linearities, which more traditional economic models often exclude.[5],[6]

Machine learning to assess inflation risk

The machine learning tool we use to assess the risks surrounding future inflation is based on a quantile regression forest (QRF) model.[7] It serves a dual purpose. First, it produces inflation forecasts. Second, it provides a comprehensive assessment of the risks surrounding a given baseline outlook, drawing on a large set of economic variables that are routinely monitored by Eurosystem inflation experts.[8] The model produces outputs based on the historical experience on which it has been trained and the most recent data available on key inflation determinants, such as wage developments and selling price expectations.

The ECB has already used the QRF model to help produce the short-term inflation forecast and the risk assessment around the baseline. Since the end of 2022, the model has become part of the broader analytical toolkit used for monetary policy preparation.[9]

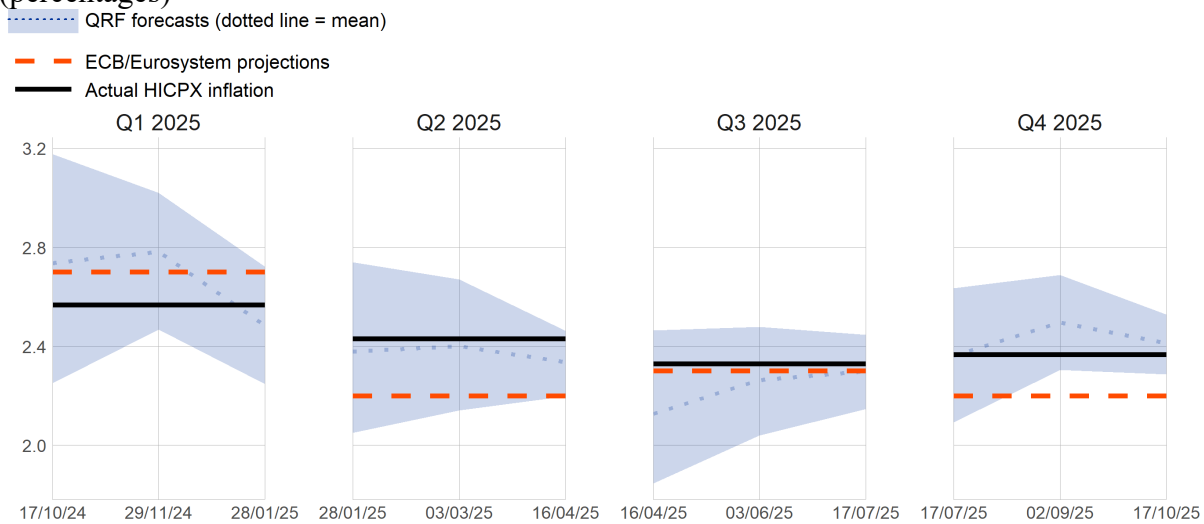
Importantly, the potential of the QRF model extends beyond forecasting, and it can help identify which factors drive inflation risks over time.[10] In 2025, for example, wages and selling price expectations were the key forces behind revisions to our projections for core inflation (HICPX).[11]

Risk signals in real time

In recent years, the insights from the QRF model have been especially relevant. In particular, the model was useful in detecting – in real time – emerging inflation risks across different components of our main inflation measure (HICP). This made a tangible difference, because the volatile environment after the pandemic made it much harder to interpret the potentially conflicting messages emerging from the large number of indicators that we traditionally monitor for our analysis. Traditional economic models struggled with these conflicting messages, but QRF helped us to make sense of them.[12]

In the following, we focus on the performance of the QRF model in the course of 2025, comparing it to the ECB/Eurosystem projections produced around the same time.[13] The shaded area in Chart 1 captures the QRF density forecasts prepared on selected dates around each ECB Governing Council meeting held during the sample period. This shows the range in which HICPX inflation is likely to fall according to the model. For example, the QRF forecasts for the first quarter of 2025 were produced on 17 October 2024 and updated on 29 November 2024 and 28 January 2025 as more information became available. The ECB/Eurosystem projections for HICPX inflation for each quarter of 2025 are shown by the dashed red line. For all quarters displayed, the projections are prepared using the information available on the mid-date of the range reported on the x-axis. Finally, the actual outcome, which is generally released soon after the end of the reference quarter, is shown by the solid black line.

Chart 1
Core inflation projections in 2025
(percentages)



Source: ECB staff calculations.

Notes: The chart compares ECB/Eurosystem projections with selected vintages of QRF density forecasts for core inflation around relevant Governing Council meetings. The QRF model is based on Lenza et al. (2025). The shaded areas represent the 16th and 84th percentiles of the QRF forecasts, the dotted line represents the mean forecast.

How big risks around the ECB/Eurosystem baseline inflation projections are can be derived from how far the QRF analysis deviates from the ECB/Eurosystem projections. For example, if the ECB/Eurosystem projection falls in the lower part of the shaded area, or even outside its lower range, the projection carries an "upside risk", i.e. the QRF analysis identifies a considerable likelihood that the final outcome for inflation will be higher than predicted. Comparing the QRF-based risk signals with how actual inflation subsequently turned out suggests that these signals were informative. For example, for the second and fourth quarters of 2025, the QRF range lies mostly or entirely above the ECB/Eurosystem projections and inflation was indeed 20 basis points above the projections. Hence, the upside risks identified by the model did materialise. At the same time, when the projections were closer to the mid-point of the range (indicating that no meaningful risks were detected by the model), the projections turned out to be more in line with the final outcomes. The shaded area of the QRF model area shrinks in the course of each quarter while, at the same time, still generally encompassing the final outcome. This shows that the QRF predictions become more precise as new information becomes available over the course of the quarter.

Implications for economic and inflation analysis and the road ahead

The examples above show that the QRF model is a helpful tool to navigate uncertain macroeconomic conditions. The ECB's experience in developing and exploiting this model points the way for future uses. First, machine learning tools can complement traditional models by providing timely information about risks, including their magnitude, orientation and determinants. Second, these new tools are valuable not only for their forecasting accuracy but also for their ability to reveal complex data patterns, such as non-linearities and sector-specific dynamics, that have become increasingly relevant for monetary policy in recent years.

The QRF framework also allows growing data volumes to be handled more efficiently. This makes it possible to assess economically interpretable risk in real time, which in turn allows changes in the baseline or the associated risks to be detected swiftly. The QRF analysis provides explanations of how key variables of interest, such as inflation or output, evolve based on the evolution of their core determinants (e.g. wages, import costs or expectations). Hence, we believe these new tools will play a growing role in forecasting, monitoring economic and inflation trends, and informing monetary policy decisions.

The views expressed in each blog entry are those of the author(s) and do not necessarily represent the views of the European Central Bank and the Eurosystem.

---

1. See point 9 of The ECB's monetary policy strategy statement (2025).
2. See also Lane, P.R. (2024), "Monetary policy under uncertainty", keynote speech at the Bank of England Watchers' Conference 2024, King's College London, London, 25 November.
3. A quantitative risk assessment survey among ECB/Eurosystem staff adds expert judgement to the risk assessment methods.
4. Machine learning is a field of study in AI.
5. The ability to capture very general economic dynamics may also reduce the need to rely on expert judgement, especially where the latter is aimed at addressing the limitations of traditional macroeconomic models.

6. For a discussion of the risk analysis based on the machine learning model in this blog, see Lane, P.R. (2024), "Disinflation in the euro area", speech at the Hutchins Center on Fiscal & Monetary Policy, Brookings Institution, Washington, DC, 8 February, in particular Chart 9; and Fleming, S., Arnold, M. and Jones, C. (2024), "Why central banks are reluctant to declare victory over inflation", *Financial Times*, 4 February, in particular the discussion of the risks identified by the "ECB's new artificial intelligence-powered model for forecasting inflation".
7. For a presentation and evaluation of the accuracy of the QRF model for euro area inflation, see Lenza, M., Moutachaker, I. and Parades, J. (2025), "Density forecasts of inflation: A quantile regression forest approach", *European Economic Review*, Vol. 178, September. See also Meinshausen, N. (2006), "Quantile Regression Forests", *Journal of Machine Learning Research*, Vol. 7, No 35, pp. 983-999.
8. The selection of potential determinants for inflation includes 60 variables capturing inflation expectations, cost pressures, real economic activity and financial conditions.
9. The QRF model has also recently been introduced as part of the toolkit for nowcasting GDP. See Eraslan, S., Fabbri, A. and Saiz, L. (2025), "Short-term forecasting of euro area economic activity in an uncertain world", *Economic Bulletin*, Issue 8, ECB.
10. See Bobeica, E., Paredes, J., Renault, T. and Rousseau, F. (2024), "Selling price expectations for services: what do they tell us about consumer price pressures?", *Economic Bulletin*, Issue 5, ECB.
11. HICPX is the Harmonised Index of Consumer Prices (HICP) excluding energy and food.
12. For more detail, see Lenza et al. (2025) who study the performance of the QRF model in detecting in real-time the risks to the inflation outlook in the ECB/Eurosystem forecasts between the fourth quarter of 2022 and the fourth quarter of 2024.
13. ECB/Eurosystem projections are produced only once each quarter, while the QRF projections are updated several times in the course of each quarter.

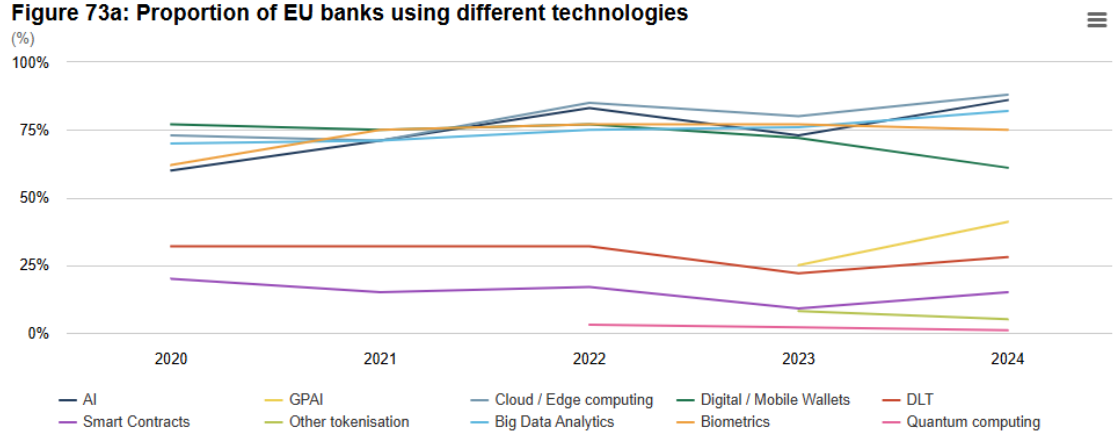
Oscar Arce
Karin Klieber
Michele Lenza
Joan Paredes

Special topic – Artificial intelligence

Broad and diverse adoption of AI in the EU/EEA banking sector

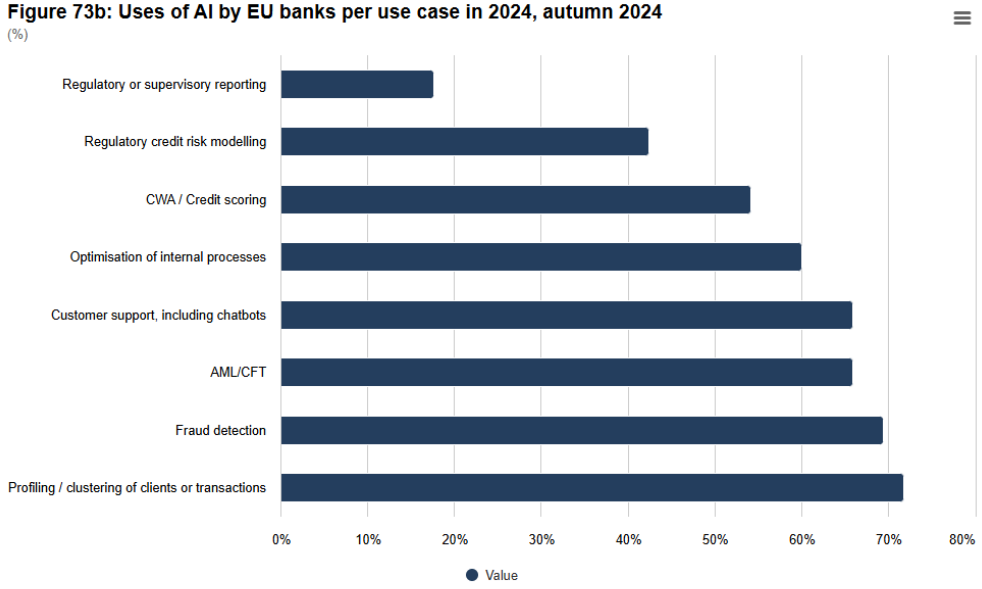
Over the past decade, the EU/EEA banking sector has undergone a profound digital transformation, embracing a broad spectrum of advanced technologies to enhance operational efficiency and customer experience. Among these technologies, AI, cloud computing, digital wallets, big data analytics, and biometrics have become increasingly prevalent, with most banks integrating them into their operations for at least the past 5 years. These technologies have played a pivotal role in reshaping banking processes, enabling more personalised services, optimising risk management, and improving decision-making processes. However, while their adoption is widespread, other innovations, such as DLT, smart contracts, and those underpinning tokenisation projects, have experienced a slower uptake (Figure 73).

Figure 73a: Proportion of EU banks using different technologies



Source: EBA Risk Assessment Questionnaire

Figure 73b: Uses of AI by EU banks per use case in 2024, autumn 2024



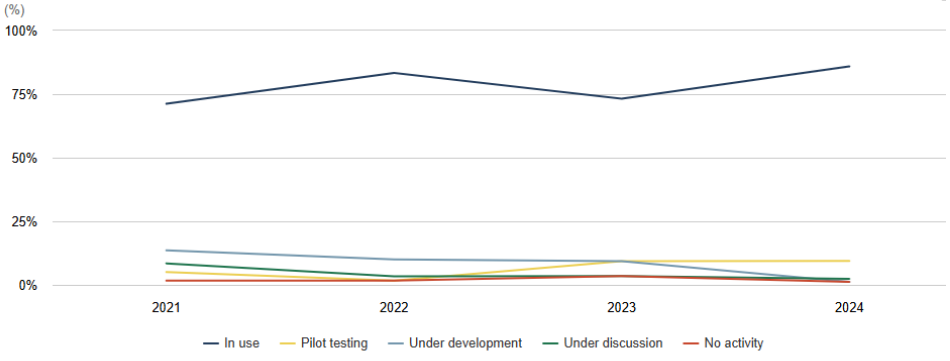
Source: EBA Risk Assessment Questionnaire

Considering the provisions of the EU AI Act[1], references to AI encompass a broad range of machine-based systems designed to operate with varying levels of autonomy. AI is generally assumed as exhibiting adaptiveness to after deployment as well as generating outputs from input data. The adoption of AI has consolidated significantly within the EU banking sector. Over the past 5 years, there has been a consistent upward trend in the deployment of AI, thanks to the potential to impact various aspects of banking. According to the EBA's RAQ (spring 2024), most EU banks are using AI methods such as regression analysis, decision trees, natural language processing and neural networks.

Regarding use cases, AI is most frequently employed in areas such as client and transaction profiling (for commercial purposes) and customer support. These use cases are consistent with trends observed in previous years' RAQs, highlighting AI's role in enhancing the accuracy and efficiency of customer segmentation and improving the responsiveness of customer service channels. Additionally, AI plays a role in fraud and AML/CFT efforts, where it is used to analyse vast amounts of data and detect patterns indicative of illicit activities. Beyond these prevalent applications, AI is also increasingly used to optimise internal processes within banks, in credit scoring and creditworthiness assessments, and in regulatory credit risk modelling. While most of EU banks are leveraging AI in these areas, there are other use cases where AI adoption is less widespread, such as supervisory reporting, monitoring conduct risk, real estate valuation or carbon footprint estimation (Figure 73).

Despite the benefits AI may bring to the EU banking sector, its adoption is accompanied by a range of challenges and risks that demand careful management and rigorous testing. Banks face competitive pressures from global financial institutions and technology companies to innovate rapidly around AI. However, many have taken a more measured approach, gradually developing their AI systems over the past years. This cautious approach is likely driven by the need to ensure compliance with existing horizontal and sectoral legislation, and a commitment to ensuring that AI is developed ethically and responsibly, aligning with regulatory standards and public expectations. Additionally, by developing AI systems in-house or retaining control over key components, banks can reduce their business and technical dependencies on third-party providers. This strategy not only enhances control over AI systems but also addresses some of the core challenges AI poses to the EU banking sector, such as data privacy, cybersecurity, and compliance with regulatory frameworks. In that sense, responses to the EBA's RAQ indicate that the recent surge in AI adoption across EU banks may be linked to a shift away from the developmental and testing phases of AI systems, as, by 2024, only a small number of banks remain in the pilot testing phase of their AI initiatives, as compared to 2022. Most of these initiatives have moved beyond pilot stages and are now fully integrating AI into their IT infrastructure (Figure 74).

Figure 74: Evolution of the adoption level of AI by EU banks



Source: EBA Risk Assessment Questionnaire

GPAI: rising interest, experimentation and early adoption

One of the most recent breakthroughs in AI is the rapidly increasing interest and experimentation with GPAI, which is also being observed in the EU banking sector. According to the EU AI Act, GPAI is AI that has the capability to serve a variety of purposes. One of the main components of GPAI systems are GPAI models (or foundation models), which are trained with large amounts of data using self-supervision at scale. One of the most popular applications of GPAI is Generative AI, which can be understood as a subset of GPAI which refers specifically to systems that can create new content such as text, images, audio or video.

Box 11: The EBA's role in GPAI

The EBA has a statutory duty to monitor and assess market developments, including financial innovation, and to achieve a coordinated approach to the regulatory and supervisory treatment of new or innovative financial activities. In accordance with this mandate and the EBA's priorities on innovative applications for 2024/2025[2], that include AI/machine learning, including GPAI, the EBA is deepening its monitoring and analysis of the uses of GPAI by EU banks. Since September 2023, the EBA has collected data on the testing and use of GPAI by EU banks via the EBA's RAQ. The results have provided the EBA with detailed insights into the levels of adoption and uses of GPAI in the EU banking sector.

Additionally, in April 2024 the EBA organised a Workshop on GPAI in the banking sector, which included the participation of a range of EU stakeholders representing banks, consumer organisations and technology providers, in addition to relevant EU and national competent authorities. The workshop aimed to promote a common understanding of the uses, risks and opportunities associated with GPAI in the banking sector. During the workshop the EBA found that the use cases of GPAI are rather limited, with activity mainly focused on testing around a small number of use cases. The EBA found that the banking sector has limited tools to fully mitigate consumer protection concerns associated with GPAI, which justifies the cautious and gradual approach to the adoption of this technology.

While GPAI activity is still largely limited to testing, experimenting and piloting, the EBA RAQ data shows an increase in its use by banks since 2023. Approximately one-third of respondent EU banks have already implemented GPAI in at least one-use case (Figure 75) highlighting the sector's growing recognition of its potential to drive efficiencies, enhance customer interactions, and generate new business insights. This rapid adoption reflects the banking sector's focus on leveraging cutting-edge technologies to remain competitive in a digital landscape. However, the increasing reliance on AI, and especially GPAI, may augment existing challenges and risks and bring new ones. This merits careful consideration and robust risk management strategies.

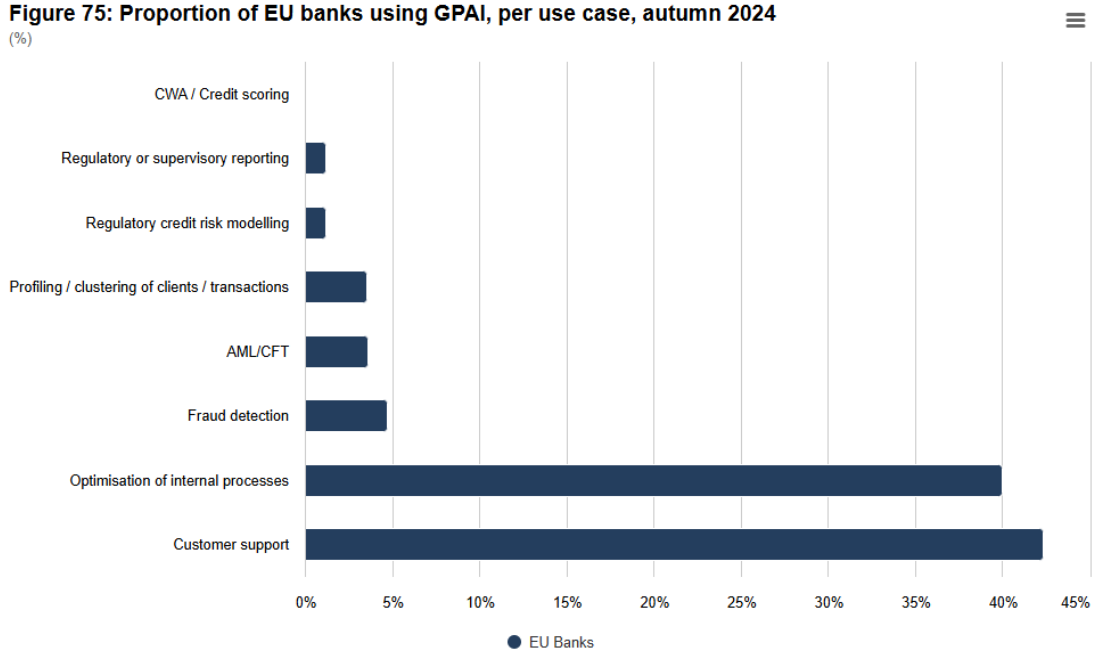
Consequently, the EBA is actively monitoring and assessing the vulnerabilities associated with a potential wider adoption of GPAI in the EU banking sector (see Box 11 above for more on the EBA's role and activities on GPAI). The EBA's objectives are to ensure that consumers continue to benefit from high-level standards of protection and that the supervisory and regulatory framework remains fit-for-purpose and harmonised across the EU, also taking account of the EU's new AI Act.

Adoption of GPAI in the EU banking sector

According to the EBA's RAQ responses, around 40% of EU banks are already using GPAI, with the adoption mainly reaching significant levels in the areas of customer support and the optimisation of internal processes (Figure 75). For instance, banks are engaging in the following non-exhaustive set of use cases:

- **Customer service (internal and external)**: GPAI can help a bank improve the resolution of customer queries via chatbot, including for employees' questions about internal policies, procedures or allowances.
- **Call centres**: GPAI can improve the transcription and summarisation of contact centre audio calls into text and the assessment of the quality and outcome of interactions.
- **Programming and coding**: GPAI can improve a bank's technology and operations units by generating new code from natural language, detecting code errors, converting code from one programming language into another or helping migrate legacy code.
- **Legal analysis**: GPAI can help a bank's legal unit improve the monitoring of legal and regulatory changes, the summarisation of court rulings, the analysis of the impact of contractual clauses and the proposal of new clauses and conditions.

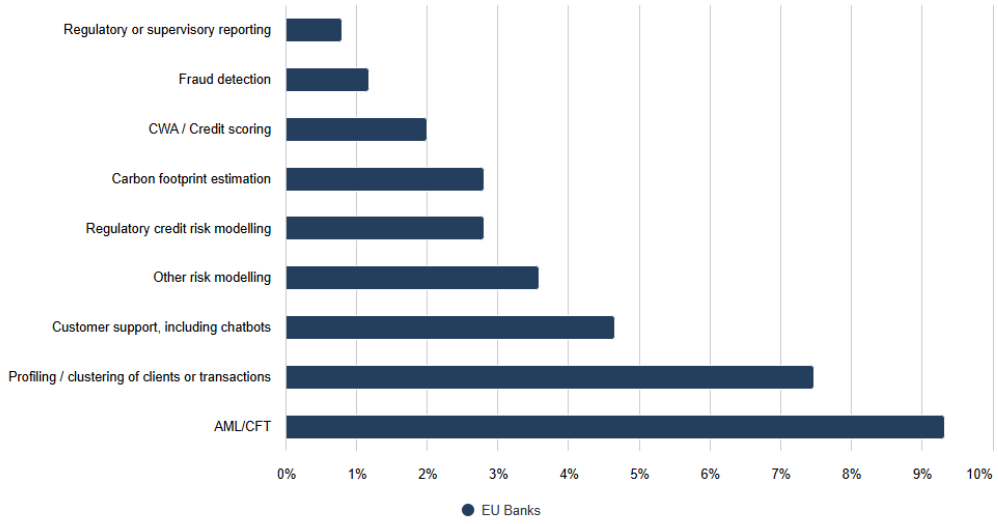
Figure 75: Proportion of EU banks using GPAI, per use case, autumn 2024



Source: EBA Risk Assessment Questionnaire

Overall, the adoption of GPAI in the EU banking sector is still at an early stage, with banks mostly testing and experimenting with GPAI via proof-of-concepts or a sandbox approach. According to the EBA's RAQ, around 10% of EU banks are already testing the use of GPAI for many other use cases, such as those related to AML/CFT and to the profiling and clustering of clients and transactions. Even in the domain of customer support, many banks are still testing and experimenting with GPAI before actively using it (Figure 76).

Figure 76: Proportion of banks testing GPAI, but still not using it in production, per use case, autumn 2024



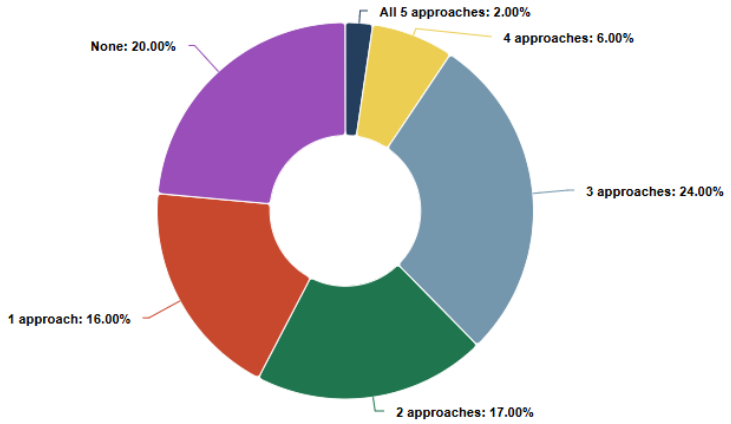
Source: EBA Risk Assessment Questionnaire

Diverse approaches for the integration of GPAI systems and models

Banks are exploring and testing various deployment approaches to integrate AI and GPAI systems or models into their technical infrastructure. Notably, banks are assessing the potential of GPAI by testing use cases, or already integrating GPAI, following different approaches, such as developing the models or systems themselves, outsourcing the development of the models or systems, or using models or systems developed by third-parties and deploying them either via Cloud application programming interfaces (APIs)[3] or 'on-premises'[4]. The choice of approach depends on factors such as scalability, cost, control over data security and compliance, need for internal resources, and internal skills and expertise. When choosing the approach, banks are also varying between relying on a single or multiple providers, or relying on proprietary or open-source models. Finally, regardless of the approach selected, to ensure and improve the performance and applicability of GPAI model for specific applications, banks are resorting to techniques such as RAG[5] or 'fine-tuning'[6].

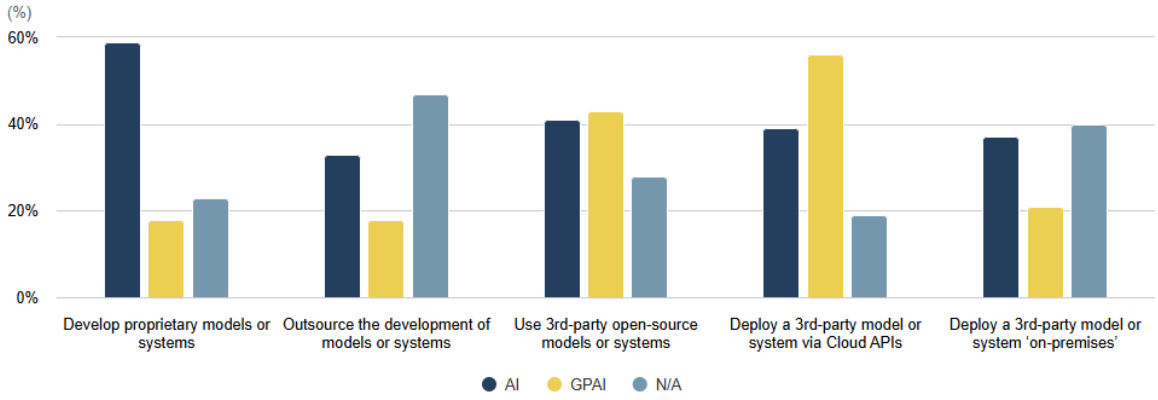
According to the EBA RAQ, EU banks most commonly deploy models and systems using between one and three different approaches. Notably, banks appear to be combining the use of third-party open-source models or systems with other methods, such as deploying third-party models via Cloud APIs or on-premises solutions and developing models in-house or outsourcing development to third parties. Generally, regardless of the combination of approaches followed, banks appear to be resorting to third parties to deploy GPAI. Most of respondent EU banks are already using third-party services, mainly via Cloud APIs offered by large model developers, due to the flexibility and scalability of doing so. A lower proportion of banks are integrating third-party GPAI systems or models 'on-premises', attracted by a higher degree of control in the hands of banks (which could benefit the scope of resources and costs dedicated to GPAI, and ultimately on the cost-effectiveness of deploying GPAI). However, many banks may face difficulties in following this deployment approach, due to limited technical skills and resources in-house. Finally, only a small number of respondent EU banks are developing proprietary GPAI models and systems, indicating the necessary financial and technical resources required as the main challenge (Figure 77).

Figure 77a: Diversity of deployment approaches by banks for GPAI. autumn 2024



Source: EBA Risk Assessment Questionnaire

Figure 77b: Deployment approaches adopted by banks to integrate or adopt AI (in general) and GPAI, autumn 2024



Source: EBA Risk Assessment Questionnaire

Hence, EU banks seem to be following a multimodal approach strategy, while assessing the effectiveness of each approach on a case-by-case basis. This may be indicative of the nascent stage of GPAI technology within the banking sector. Each approach entails different opportunities and risks, and decisions are influenced by cost, security, skills availability in-house (and via third parties), quality of data available and data privacy considerations. As banks are still primarily in the testing and piloting phase, experimentation allows banks to assess the strengths and limitations of various approaches.

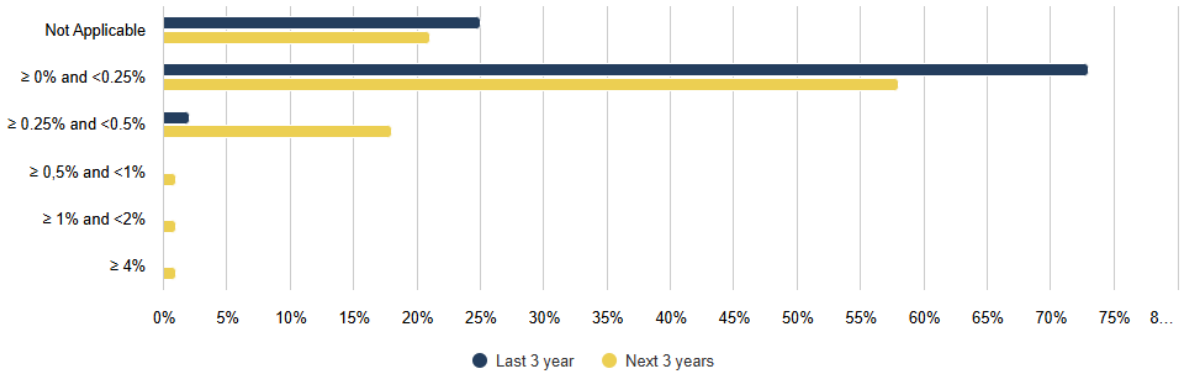
Drivers and obstacles to GPAI adoption in the EU banking sector

To-date, the EBA has observed that GPAI applications currently being tested and experimented with by banks focus on areas where the technology can bring operational efficiencies, either via increasing productivity, lowering costs, or supporting and improving the speed of back-office processes (such as, coding, programming, auditing or fraud detection). Banks appear to be attracted to GPAI adoption by efficiency gains in areas such as content intelligence, content generation, customer engagement or code generation.

In addition, some EU banks are engaging with GPAI as part of wider initiatives to foster an innovative culture and position their entity as innovation-friendly towards customers and other businesses. In this sense, banks across the EU are showing a considerable interest in exploring

the potential of GPAI in financial terms as well. According to the EBA RAQ data, while more than half of the respondent banks plan to invest between 0% and 0.25% of their equity in GPAI, around 21% of respondent EU banks have indicated plans to invest above 0.25% of equity, with one bank in the sample planning a substantial investment exceeding 4% of equity into GPAI (Figure 78). These investment figures may suggest that EU banks are interested in the potential of GPAI but are approaching it with a measured level of financial commitment, likely reflecting the early stages of its adoption and the need to carefully assess the return on investment[7].

Figure 78: Proportion of banks investing or planning to invest in GPAI (investment levels as % of equity), autumn 2024



Source: EBA Risk Assessment Questionnaire

Despite the abovementioned drivers, the EBA observes that currently EU banks do not consider GPAI is sufficiently robust to be used in other use cases. However, the EBA notes that some are researching around the potential of GPAI in other areas such as AML/CFT, the profiling and clustering of clients or risk modelling. In doing so, some EU banks are experimenting with GPAI in those use cases. Consequently, the EBA will keep monitoring those uses via surveys, workshops and desk-based research, as research and experience of the banking sector with GPAI improves.

Potential risks associated with the use of GPAI in the banking sector

In terms of the risks associated with GPAI, as compared to 'other AI' or 'traditional AI', based on engagement with competent authorities and industry, the EBA has identified potential risks regarding the following:

- **Explainability**. Several parameters are used to train GPAI models and randomness is a key feature of outputs. As a result, GPAI models and systems can be highly complex and opaque. Consequently, explainability techniques used for 'traditional AI', which are limited in efficiency, are not necessarily valid for GPAI. Besides, GPAI-specific explainability techniques (such as, RAG techniques or asking a GPAI model to point back to the source of an output) are not sufficient to comply with explainability expectations on consumer-facing interfaces. The explainability challenges therefore appear to be difficult to tackle considering the existing research and technical developments.
- **Reliability**. In addition to the abovementioned explainability challenges, most of the state-of-the-art GPAI models suffer from 'hallucinations' - i.e. output generated by GPAI which contains incorrect or misleading information presented as fact to the end-

user, for which researchers have not yet found a solution. Consequently, the outputs of GPAI systems may face reliability issues, which may justify the cautious approach that banks are adopting towards the adoption of GPAI.

- **Transparency**. Transparency around the capabilities and limitations of GPAI models is particularly important for the banking sector, considering the requirements that banks face regarding model validation, privacy, security, accessibility or consumer protection. While GPAI model providers often develop GPAI-specific technical documentation (e.g. 'model cards'), there are still some concerns observed related to privacy and intellectual property issues.
- **ICT risks**. GPAI models are often developed and provided by third-party technical providers. Therefore, the integration of GPAI into a credit institution's technical architecture may enhance challenges to operational resilience, as well as security and privacy concerns. While banks face similar challenges when integrating other technical service providers, the complexity of GPAI applications and the limited explainability of their outputs may increase ICT risks.
- **Data governance**. GPAI models rely on training data that may have limitations in quality, reliability and privacy. Due to the extensive size of datasets used, GPAI could present challenges in ensuring sound data governance for banks.
- **Access to necessary skills**. The use of GPAI introduces challenges that require banks to adopt a 'human-in-the-loop' approach, involving human intervention throughout the application's lifecycle. However, the rapid development in GPAI research and the limited availability of expertise in this area can make it difficult for banks to access the necessary skills and talent. This has led to an effort in upskilling and re-skilling efforts, especially among larger banks. Smaller banks, however, might face difficulties competing for the required skills and talent.
- **Other concerns**: GPAI also presents challenges such as environmental impacts from the development of GPAI models (i.e. water and electricity consumption) or concerns around competition and market concentration due to the substantial investments needed to compete in the development of state-of-the-art GPAI models and the limited number of technical services providers in the market.

Box 12: Consumer protection challenges associated to GPAI with the EU banking sector

Based on engagement with competent authorities, market stakeholders and consumer organisations, the EBA has identified that the following consumer protection issues could arise if banks adopt GPAI in consumer-facing use cases:

Accountability. The complexity and limited transparency associated with GPAI could raise challenges in terms of ensuring that providers of GPAI systems and models to banks are accountable for inaccurate or inappropriate outcomes, such as inaccurate, misleading or false information or advice to bank customers. However, as banks are accountable for inaccurate, misleading, or false information provided to consumers according to general financial protection legislation, consumers may seek compensation for harms through banks.

Bias, discrimination and financial exclusion. Training data used by GPAI model developers can, without mitigation, exacerbate discrimination and bias against minority or misrepresented groups, leading to risks of financial exclusion. Due to the randomness associated with GPAI outputs, banks deploying GPAI in customer-facing applications may face limited capabilities for understanding and explaining the logic behind biased outputs. Additionally, for instance, consumer support for speakers of minority languages may be impacted if such support

gradually shifts from human to GPAI-powered virtual assistants, since GPAI models are mainly trained on the largest languages.

Transparency. The opacity of GPAI applications implies that informing consumers about the use of GPAI in consumer-facing interfaces may not be sufficient for raising their awareness. Therefore, banks might need to explore the use of other means (e.g. application-specific disclaimers) to provide more complete information to consumers.

Data security. The improved technical abilities of attackers (including new types of cyber attacks, prompt-related attacks, misinformation campaigns, data poisoning attacks, and data exfiltration techniques) can raise data security concerns for consumers. Use of GPAI can also contribute to data security risks for consumers via information leakage or inappropriate or non-factual responses. However, thanks to GPAI's potential benefits in areas such as programming script analysis or malware detection and investigation, existing research[8] suggests that there is no substantial evidence yet suggesting that GPAI can automate sophisticated cybersecurity tasks which could tip the balance between cyberattackers and defenders in favour of attackers.

Other concerns. As GPAI can potentially change the threat landscape for banking consumers by boosting cyberattackers' technical skills, aiding social engineering or phishing attacks, and speeding up manipulation techniques through consumer personification and hyper-realistic scams, GPAI could introduce challenges in consumer experiences.

In view of these potential risks, EU banks appear to be adopting a risk-based and graduated approach to GPAI and are focusing on building out guardrails, controls and ensuring human intervention during the early adoption of GPAI. As a consequence, higher risk use cases are being tested by banks only after they have reached an advanced understanding of GPAI models, deployment methods and their potential effects and necessary mitigants.

Abbreviations and acronyms

[1] The EU AI Act it entered into force in August 2024. However, different elements of the Regulation will be applicable in a phase approach, with the full application expected by 2 August 2026. See: http://data.europa.eu/eli/reg/2024/1689/oj

[2] EBA Digital Finance

[3] The use of cloud-based APIs enables banks to access and integrate GPAI models into their systems without the need for extensive in-house infrastructure.

[4] The 'on-premises' solution involves installing and running GPAI models directly on a bank's own servers and infrastructure, providing greater control over data security and compliance, but requiring significant internal resources for maintenance and scalability.

[5] RAG techniques aim to optimise the output of GPAI models with facts retrieved from authoritative sources distinct from training datasets, thus extending their capabilities to specific domains or an organisation's internal knowledge base, without the need to retrain the model.

[6] Fine-tuning is a technique in which pre-trained GPAI models are customised to perform specific tasks or behaviours by using examples and instructions.

[7] However, the EBA notes that a quarter of the respondent banks have indicated plans to increase their investments in GPAI over the coming years. Nevertheless, most banks either do not plan to increase their investments in GPAI or were unable or unwilling to provide a response to this question. This may reflect ongoing uncertainties about the long-term strategic value of GPAI or signal that banks are still in the process of determining the role this technology will play in their broader digital transformation strategies

[8] See the Interim International Scientific Report on the Safety of Advanced AI, published in May 2024: https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai

7

L’intelligence artificielle : bénédiction ou malédiction pour la transformation du secteur financier ?

Discours de Denis Beau
Premier sous-gouverneur de la Banque de France
Singapour, 8 novembre 2024

Mesdames et Messieurs,

Commençons par l’évidence : l’intelligence artificielle (IA) suscite actuellement un fort engouement. Les possibilités offertes par cet ensemble de technologies semblent en effet immenses. L’IA est à la base de tels progrès que, cette année, le prix Nobel de physique a été attribué à deux pionniers des réseaux de neurones artificiels, tandis que le prix Nobel de chimie récompensait une application de l’IA pour comprendre la structure des protéines.

Dans le secteur financier, l’IA est également de plus en plus utilisée, par exemple pour évaluer le risque de crédit, fixer les taux d’assurance ou estimer la volatilité des actifs. Cette technologie est aujourd’hui **le principal moteur de la transformation numérique du secteur financier**. Et cette transformation ne peut évidemment pas laisser insensible le banquier central et superviseur du secteur financier que je suis. Elle pose en effet la question des **opportunités** et des **risques** de ces mutations pour secteur financier. Et elle interroge aussi les banques centrales et les superviseurs financiers dans leur propre fonctionnement et dans leur **maîtrise de ces nouvelles technologies** : comment surveillera-t-on demain l’utilisation de l’IA par les acteurs financiers, et quelles sont les possibilités offertes par ces technologies pour nos propres métiers ? Permettez-moi d'aborder brièvement l’une et l’autre de ces questions.

I/ En ce qui concerne la première question des opportunités et des risques, établissons d’abord ce premier constat pour mettre les choses dans la bonne perspective : **l’IA, couplée à la data science, est un puissant moteur de transformation du secteur financier**.

1/ Nos observations montrent que **l’IA est utilisée de façon croissante par les institutions financières**, dans tous les segments de la chaine de valeur. Celles-ci s’en servent en particulier pour améliorer « l’expérience utilisateur », mais aussi pour automatiser et optimiser nombre de processus internes. L’IA est également utilisée pour le contrôle et la maîtrise des risques, comme en témoigne son succès dans les cas d’usage liés à la lutte contre la fraude ou contre le blanchiment et le financement du terrorisme (LCB-FT).

L’arrivée de l’IA générative, depuis maintenant deux ans, a constitué un puissant facteur d’accélération des tendances déjà engagées. Elle a en effet conduit à une **révolution de l’accessibilité des technologies d’IA** : la possibilité d’interagir en langage naturel avec des algorithmes -via les grands modèles de langage (*large language models* ou LLM)- facilite considérablement l’adoption des nouvelles technologies. Elle accélère aussi la **dynamique de l’innovation** au sein des entreprises, car l’écriture de code n’est plus réservée aux informaticiens. L’IA pourrait ainsi avoir un impact puissant sur la productivité : en France, la Commission de l’IA, présidée par Anne Bouverot et Philippe Aghion, a évalué à environ 1 % par an le surplus de croissance que le déploiement de l’IA pourrait générer dans notre pays au cours des dix prochaines années.

Cette révolution technologique permet de développer de **nombreux cas d'usage**. Les *chatbots* nourris à l'IA générative ont par exemple la capacité d'apporter une réponse personnalisée et plus adaptée à la situation de l'utilisateur. En assurance, le traitement des images satellite permet dans certains cas d'évaluer directement les dommages causés à un bâtiment, puis de gérer automatiquement l'indemnisation des clients. Et ce n'est que le début : nous sommes probablement loin d'avoir exploré toutes les possibilités offertes par ces technologies.

Bien maîtrisée, l'IA est donc susceptible **d'accroître l'efficacité des institutions financières et de contribuer à augmenter leurs revenus**, jouant ainsi positivement sur leur rentabilité -élément central de leur solidité-, y compris en leur offrant des solutions de contrôle des risques.

2/ **Pour autant, la médaille a un revers, et la puissance des solutions développées s'accompagne de risques significatifs**. Je voudrais évoquer ici deux d'entre eux.

D'abord, le risque de mauvaise utilisation de ces technologies. La **complexité** et la **nouveauté** de certaines modélisations peuvent en effet donner lieu à davantage d'erreurs, soit dans la conception soit dans l'utilisation des systèmes. Il s'agit d'un risque pour la clientèle, mais aussi pour la santé financière des établissements, car un modèle mal calibré pourra engendrer des pertes systématiques. **Deux éléments renforcent ces risques**. D'abord, **l'ajustement en temps réel des paramètres** de certains modèles, qui fait leur force, peut aussi se traduire par une **dérive rapide**. Ensuite, certains systèmes d'IA sont **particulièrement opaques**, générant un phénomène de « **boîte noire** ». C'est bien sûr un enjeu de protection de la clientèle, car un client doit pouvoir comprendre une décision automatisée prise à son endroit. Mais c'est aussi un enjeu de gouvernance : un établissement qui comprend mal les décisions que prennent ses systèmes d'IA ne peut prétendre en **maîtriser les risques**.

Je voudrais mentionner un second risque, et non des moindres : le **risque cyber**. Celui-ci est devenu au cours des dernières années le premier risque opérationnel dans le secteur financier. **Or l'IA est un facteur d'amplification de ce risque**, et notamment parce que la technologie accroît fortement la dangerosité des attaquants : assistants en écriture de code détournés pour concevoir des logiciels malveillants, voix synthétiques facilitant l'usurpation d'identité… La liste des menaces est longue. C'est d'ailleurs l'une des raisons d'être du règlement européen DORA, qui entrera en vigueur à partir de janvier 2025.

Toutefois, la technologie pourra aussi être mobilisée pour contrer ces attaques. **L'IA constitue ainsi un outil à double tranchant**, comme l'illustre parfaitement le domaine des paiements. Dans les paiements, l'IA peut considérablement faciliter les escroqueries, par exemple au moyen de *deepfakes*. Mais cette technologie va probablement aussi devenir un **allié essentiel dans la lutte contre la fraude**, en permettant d'identifier plus efficacement et plus rapidement les schémas de fraude. D'une certaine manière, il s'agit d'une histoire du type « Attrape-moi si tu peux » -pour faire référence à un film bien connu des experts des paiements-, où les fraudeurs et les autorités recourent à des tactiques de plus en plus sophistiquées pour se déjouer les uns les autres.

Dans le cadre de son mandat légal visant à garantir la sécurité des instruments de paiement scripturaux, la Banque de France est le fer de lance des efforts visant à évaluer les **implications de l'IA pour la sécurité des paiements**, au travers de l'Observatoire de la sécurité des moyens de paiement (OSMP), que j'ai le privilège de présider. Des **outils de notation avancés** sont ainsi utilisés depuis des années pour réduire les fraudes sur les paiements par carte. Ces outils continuent d'être améliorés, en s'appuyant notamment sur des protocoles d'authentification forte tels que 3-D Secure. Alors que les paiements instantanés se développent rapidement, **je crois que le moment est venu d'enrichir ces outils en utilisant l'IA pour sécuriser d'autres canaux de paiement tels que les virements, les prélèvements automatiques et les transferts de fonds**.

II/ Mais la révolution technologique engendrée par l'IA constitue également un moteur de transformation pour les autorités financières, autour d'un double mouvement : **celles-ci doivent à la fois se préparer à superviser l'utilisation de l'IA par le secteur financier, et utiliser les nouvelles technologies pour améliorer leur efficacité et développer de nouvelles capacités**.

1/ Face aux risques de l'IA, et afin de permettre au secteur financier de tirer pleinement parti de ces opportunités, **il nous faut construire une régulation efficace**. Le mouvement d'encadrement réglementaire a déjà commencé : avec son **règlement sur l'IA (« AI Act »)**, entré en vigueur cet été, l'Union européenne s'est dotée du **premier instrument législatif au monde** et a posé les bases d'une « IA de confiance ». Pour ce faire, le règlement distingue plusieurs niveaux de risques, au sein desquels les **« risques élevés » -**qui constituent le cœur du texte- vont **concerner le secteur financier à au moins deux titres** : l'évaluation de la solvabilité pour l'octroi de crédit à des personnes physiques ; l'évaluation des risques et la tarification en assurance santé et en assurance-vie.

Pour ces cas d'usage du secteur financier, **le règlement confie le rôle « d'autorité de surveillance du marché » aux superviseurs financiers nationaux -**et donc en France à l'ACPR, l'autorité de contrôle prudentiel et de résolution-. **C'est un choix judicieux !** Il permettra d'articuler au mieux la mise en œuvre de ce texte avec la réglementation sectorielle existante, avec l'aide -le moment venu- **d'orientations des autorités européennes de supervision**.

L'ACPR se tient prête à exercer le nouveau rôle que le règlement européen va lui confier. Elle le fera en se fondant sur une approche « basée sur les risques », et en utilisant au maximum les synergies existantes avec le contrôle prudentiel, afin de limiter la charge supplémentaire pour les établissements comme pour elle-même.

Je ne minimise cependant pas le défi que la mise en œuvre de l'AI Act représente : au-delà des aspects organisationnels, il nous faudra élaborer une méthodologie de l'audit de l'IA, ce qui constitue à n'en pas douter un saut qualitatif dans nos méthodes de travail. **Ce défi, nous l'accueillons non pas comme une contrainte, mais comme une opportunité** : partout dans le monde, en effet, qu'une réglementation spécifique ait été adoptée ou non, les superviseurs financiers doivent développer de nouvelles capacités afin de répondre à l'utilisation croissante de l'IA par les acteurs financiers. Et, de fait, nous entendons construire des **coopérations efficaces** avec d'autres superviseurs, et notamment au niveau européen et international.

2/ **Nous voulons également utiliser les technologies d'IA pour nos propres besoins**. Nous entendons ainsi améliorer notre **efficacité opérationnelle**, tout en donnant aux agents de la Banque de France et de l'ACPR de **nouvelles capacités**. Nous entendons en effet construire un cadre d'IA de confiance, dans lequel la technologie vient en *complément* des humains et non en *remplacement*.

L'investissement dans l'IA fait ainsi l'objet d'une **action spécifique de notre plan stratégique**, suivie au plus haut niveau. Cette action a **deux ambitions principales** : offrir une gamme de services génériques en ligne à tous nos agents -avec par exemple des capacités de recherche augmentées, ou encore des fonctionnalités d'analyse et de synthèse de documents-, et fournir des assistants IA pour répondre aux besoins métier plus spécifiques.

Parce que **l'apprentissage par la pratique est le meilleur chemin pour progresser**, nous commencerons par 5 cas d'usage prioritaires à court terme. Pour illustrer mon propos, j'en citerai trois : un assistant conversationnel interne pour répondre aux interrogations des agents sur les sujets RH, un outil visant à détecter les opérations atypiques pour la LCB-FT -notamment dans les opérations du Trésor français gérées par la Banque de France-, ou encore un outil de cartographie des produits financiers structurés pour l'ACPR.

Un dernier point pour conclure : **le déploiement de l'IA à la Banque de France et à l'ACPR se fait de manière raisonnée et maîtrisée**. Ainsi, nous ne recourons à des infrastructures de cloud public que lorsque les traitements portent sur des données non sensibles. Plus généralement, nous nous assurons d'avoir une **compréhension approfondie** des modèles et de leurs résultats afin de conserver le contrôle total sur nos activités. Et nous déployons et intégrons l'IA **graduellement**, afin de laisser toute sa place à la formation des agents, au retour d'expérience et à l'amélioration continue.

Je vous remercie pour votre attention.

8

Deepfakes : l'image de la Banque de France détournée pour piéger les entreprises

La Banque de France et l'autorité de régulation (ACPR) alertent sur une vague de fraudes sophistiquées utilisant des trucages numériques. En usurpant l'image de leurs dirigeants via l'intelligence artificielle, les escrocs tentent de valider des transactions illicites.

Les informations à retenir

Quelle est la nouvelle menace de fraude identifiée par la Banque de France ?

- Des escrocs utilisent l'IA pour créer des deepfakes (voix et visages) du gouverneur François Villeroy de Galhau et de la secrétaire générale de l'ACPR.
- L'objectif est de simuler une autorité institutionnelle pour exiger des virements urgents sous couvert de frais de certification ou de taxes.
- La Banque de France rappelle qu'elle ne communique jamais par messagerie instantanée pour des opérations de transfert de fonds.

L'alerte est tombée lundi en fin de journée avec la solennité des communiqués de crise que l'on réserve d'ordinaire aux faillites bancaires ou aux krachs. La Banque de France et l'Autorité de contrôle prudentiel et de résolution (ACPR) dénoncent une offensive d'escrocs d'un genre nouveau : l'usurpation d'identité par intelligence artificielle générative.

Le cœur de cette attaque réside dans une mise en scène millimétrée du prestige institutionnel. Selon le communiqué officiel, les fraudeurs ne se contentent plus de détourner des logos ; ils s'approprient les visages et les voix des plus hauts responsables de la finance française. François Villeroy de Galhau, gouverneur de la Banque de France, et Emmanuelle Assouan, secrétaire générale de l'ACPR, voient leurs identités numériques détournées pour cautionner des actions frauduleuses lors de visioconférences ou de messages diffusés sur les réseaux sociaux.

Or, recevoir une consigne en apparence signée par le régulateur national paralyse souvent l'esprit critique. Cela s'applique aussi bien aux responsables financiers qu'aux chefs d'entreprise en situation de vulnérabilité. Les fraudeurs utilisent ces trucages numériques (ou *deepfakes*) pour simuler une autorité naturelle, dirigeant leurs cibles vers des interfaces frauduleuses sous couvert de procédures réglementaires d'urgence ou de déblocages de fonds complexes.

L'ingénierie sociale à l'heure de la synthèse totale

Le mode opératoire décrit par les autorités est d'une efficacité redoutable, car il s'appuie sur une vérité psychologique simple : on ne discute pas les ordres de l'arbitre. Les escrocs ciblent des situations d'attente financière tendue : une succession en cours, des fonds placés sur des comptes d'épargne oubliés ou des héritages bloqués par une administration imaginaire.

Le déroulement suit un scénario bien établi. L'interlocuteur se présente comme un dirigeant de la Banque de France, souvent spécialisé dans la supervision des moyens de paiement pour accroître sa crédibilité technique. Pour lever les doutes, un lien vers une vidéo ou un audio est

fourni. La voix est identique à l'original grâce à l'apprentissage automatique (machine learning) nourri par les multiples interventions publiques des dirigeants sur les plateaux de télévision.

Une fois le contact visuel ou sonore établi, les fraudeurs expliquent que pour débloquer ces sommes, des « frais de certification » ou des « taxes de transfert » doivent être versés immédiatement. Une fois le virement effectué, les interlocuteurs s'évaporent et les fonds sont blanchis via des réseaux internationaux de cryptomonnaies, rendant toute récupération quasi illusoire.

Pourquoi les systèmes de défense vacillent

Pendant des décennies, la sécurité des virements reposait sur la détection d'anomalies visibles : un accent étranger, des fautes de syntaxe dans un mail ou une adresse d'expédition suspecte. L'intelligence artificielle générative a balayé ces garde-fous en quelques mois. Un logiciel de clonage vocal peut désormais reproduire une voix humaine avec seulement trente secondes d'échantillon sonore.

Le résultat est une voix capable d'exprimer la nuance, l'urgence ou la solennité, rendant la détection humaine impossible lors d'un simple appel. Au-delà de la technologie, c'est l'exploitation du levier de l'autorité institutionnelle qui fait le succès de ces fraudes. La Banque de France incarne la stabilité du système ; obéir à ses directives semble être un acte de conformité plutôt qu'un risque financier.

Le dernier rempart : la résistance humaine

Face à cette menace, la réponse ne peut pas être uniquement logicielle. Elle doit être organisationnelle. D'abord, il faut instaurer une culture du doute systématique : plus l'interlocuteur semble important et la demande pressante, plus le contrôle doit être rigoureux.

Une astuce technique simple permet de démasquer les *deepfakes* actuels lors d'une visio : demander à l'interlocuteur de se tourner de profil ou de passer sa main devant son visage. Ces mouvements perturbent les algorithmes de génération d'image en temps réel, créant des saccades ou des artefacts visuels qui révèlent la supercherie. De même, la Banque de France ne communique jamais par messagerie instantanée type WhatsApp ou Telegram pour des opérations de transfert de fonds.

Vers une crise de confiance numérique ?

Cette alerte pose une question fondamentale sur la survie de la confiance dans l'espace numérique souverain. Si l'image d'un régulateur peut être usurpée avec une telle facilité, c'est l'ensemble de la chaîne du paiement qui est fragilisé. À l'avenir, la sécurité passera sans doute par des protocoles d'authentification physique ou des signatures numériques cryptographiques impossibles à imiter pour une IA. En attendant ces solutions, la vigilance humaine et le contre-appel systématique vers les numéros officiels restent les seules armes efficaces contre cette nouvelle forme d'escroquerie.

9

The Financial Stability Implications of Artificial Intelligence

The rapid adoption of AI in finance means that authorities should address information gaps for monitoring, assess the adequacy of current policy frameworks and enhance supervisory and regulatory capabilities.

This report revisits the 2017 FSB report on AI and machine learning in financial services by taking stock of recent advancements, exploring use cases in the financial sector and drivers of adoption, as well as new potential benefits and AI-related financial sector vulnerabilities.

In the past few years, technological advancements and increased computational power have led to an uptake in AI adoption by financial firms and supervisors. AI offers benefits such as increased operational efficiency, regulatory compliance, financial product customisation and advanced analytics. With the advent of generative AI (GenAI) and large language models, the range of use cases has become more diverse.

While AI offers benefits like improved operational efficiency, regulatory compliance, personalised financial products, and advanced data analytics, it may also potentially amplify certain financial sector vulnerabilities. AI-related vulnerabilities that stand out for their potential to increase systemic risk include: (i) third-party dependencies and service provider concentration; (ii) market correlations; (iii) cyber risks; and (iv) model risk, data quality and governance. GenAI also increases the potential for financial fraud and disinformation in financial markets. Misaligned AI systems that are not calibrated to operate within legal, regulatory, and ethical boundaries can also engage in behaviour that harms financial stability. And from a longer-term perspective, AI uptake could also drive changes in market structure, macroeconomic conditions and energy use that could have implications for financial markets and institutions.

While existing regulatory and supervisory frameworks address many of the vulnerabilities associated with AI adoption, more work may be needed to ensure that these frameworks are sufficient. The report calls for national financial authorities and international bodies to enhance monitoring of AI developments, assess whether financial policy frameworks are adequate, and enhance their regulatory and supervisory capabilities including by using AI-powered tools.

10

Financial Stability Risks Mount as Artificial Intelligence Fuels Cyberattacks

Resilience, supervision, and international coordination are essential to safeguarding global financial markets as new AI tools enable attackers

Artificial intelligence is transforming how the financial system copes with vulnerabilities and reacts to incidents. Yet it is also amplifying cyber threats that can undermine financial stability when the offensive capabilities of intruders outpace defenses.

IMF analysis suggests that extreme cyber-incident losses could trigger funding strains, raise solvency concerns, and disrupt broader markets.

The financial system relies on shared digital infrastructure that's highly interconnected, including software, cloud services, and networks for payments and other data. Advanced AI models can dramatically reduce the time and cost needed to identify and exploit vulnerabilities, raising the likelihood of simultaneously discovering and targeting weaknesses in widely used systems. As a result, cyber risk is increasingly about correlated failures that could disrupt financial intermediation, payments, and confidence at the systemic level.

Anthropic's recent controlled release of its Claude Mythos Preview, an advanced AI model with exceptional cyber capabilities, underscored how quickly risks are increasing. Mythos could find and exploit vulnerabilities in every major operating system and web browser—even when used by non-experts. This foreshadows how fast-moving, AI-driven cyber risks could destabilize the financial system if not managed carefully, and why authorities must focus on building resilience through supervision and coordination—rather than treating these developments as purely technical or operational issues.

On the other hand, OpenAI's specialized, restricted cyber version of GPT-5.5 assumes vulnerabilities and attacks will grow, and emphasizes equipping defenders more quickly and at scale, under appropriate governance and trusted access models.

Advances change risk equation

Models such as Mythos illustrate the nature of the challenge because they amplify existing cyberattack techniques by operating at machine speed. Attackers have the advantage over defenders because discovering and exploiting vulnerabilities can occur faster than patching and remediation. In a financial system built on common software and shared service providers, this can create simultaneous vulnerabilities across many institutions.

For now, some mitigating factors remain. Advanced AI cyber capabilities are not yet widely available, and closed, industry-specific financial software is harder to target than open-source infrastructure. But these buffers are likely to erode quickly as model training expands, capabilities diffuse, and leaks occur. Temporary containment is unlikely to substitute for durable defenses.

Financial stability implications

The new AI-enabled cyber tools focus the discussion on financial stability:

- Risks are systemic. Attacks become more dangerous when discovery and exploitation scale rapidly, with implications for financial stability.
- Risks cut across sectors. The financial sector shares digital foundations with energy, telecommunications, and public services. That means AI-assisted attacks can propagate across sectors that rely on the same infrastructure.
- AI may further concentrate risk and failures with one vulnerability rippling across many institutions. Reliance on a small number of software platforms, cloud providers, or AI models increases the impact of any single exploited weakness.

These features elevate cyber risk to a potential macro-financial shock. Confidence effects, payment disruptions, liquidity strains, and fire-sale dynamics could follow if multiple institutions are affected simultaneously. For financial authorities, the question is whether the system is prepared to absorb cyber incidents without destabilizing core financial functions.

AI in cyber defense

AI is also a critical part of the solution. When attackers operate at machine speed, defenders must do the same. Financial institutions increasingly use AI-supported tools to detect threats, prevent fraud, identify vulnerabilities, and respond to incidents.

AI also can help reduce vulnerabilities at the development stage rather than patching them after release. For widely used financial infrastructure, these gains can meaningfully reduce systemic exposure. But these benefits will materialize only if institutions invest in integration, governance, and human oversight—areas that supervisors increasingly need to assess. This also includes business continuity and disaster recovery, cyber and quality assurance programs, and good cyber hygiene practices.

Resilience-first policy framework

AI-driven cyber risk demands a policy response that treats cybersecurity as a core financial stability issue. Existing measures remain relevant, but they must be expanded and sharpened for a world of faster, automated, and increasingly sophisticated attacks. Policymakers should prioritize robust resilience standards, supervision focused on systemic transmission channels, and close public-private collaboration on threat intelligence and incident response.

Defenses will inevitably be breached, so resilience must also be a priority, specifically to limit how far incidents spread and ensure rapid recovery. Controls to stop the spread of attacks can prevent local breaches from escalating into system-wide disruptions. These measures are often costly and complex, but they are among the most effective tools for containing AI-enabled attacks.

From a supervisory perspective, this underscores the need to focus not only on prevention, but on response, recovery, and continuity of critical functions. Cyber stress testing, scenario analysis, and board-level oversight of cyber risk are becoming indispensable components of financial stability frameworks.

International cooperation is vital

The Mythos episode also highlights governance challenges. Cyber risk does not respect borders. As AI capabilities spread across countries, inconsistent oversight could weaken a globally interconnected system.

Emerging and developing economies, which often have more severe resource constraints, may be disproportionately exposed to attackers targeting regions with weaker defenses. That's why stronger international coordination, more information sharing, and expanded capacity development are critical to preserving global financial stability.

As AI reshapes the cyber landscape, the central question for authorities is whether the financial system can continue to function under severe stress. Answering that question requires putting systemic risk—and the tools to manage it—at the center of the AI-cyber conversation.

Tobias Adrian
Tamas Gaidosch
Rangachary Ravikumar

3. Les enjeux de l'intelligence artificielle pour le système financier

Par Claire Brousse, Olivier Fliche, Jules Joyez et Julien Uri

Les technologies basées sur l'intelligence artificielle offrent de nombreuses opportunités pour le secteur financier, avec des applications notamment pour le service clients, l'optimisation des processus, la gestion des risques ou les activités de marché. Néanmoins, par sa nature, l'intelligence artificielle présente des enjeux de gestion des données, de modélisation ou de gouvernance. Elle pourrait accentuer des risques pour la stabilité financière[125], en particulier les risques de procyclicité et de volatilité des marchés, le risque d'une concentration des marchés et le risque de cyberattaques. En outre, des risques de mauvais usages, tels que la manipulation de l'information, pourraient déstabiliser le système financier. Les impacts potentiels de l'intelligence artificielle restent difficiles à estimer, car ils dépendront de l'ampleur et des modalités de son adoption. Néanmoins, une gouvernance et un cadre de gestion des risques adaptés aux enjeux de l'intelligence artificielle sont nécessaires pour en maîtriser les risques. La législation européenne sur l'intelligence artificielle, adoptée en mai 2024, est conçue spécifiquement pour favoriser une intelligence artificielle digne de confiance.

3.1. Le développement de l'intelligence artificielle présente des opportunités pour le système financier et l'économie

Les technologies d'intelligence artificielle ont connu une accélération spectaculaire au cours des dernières années

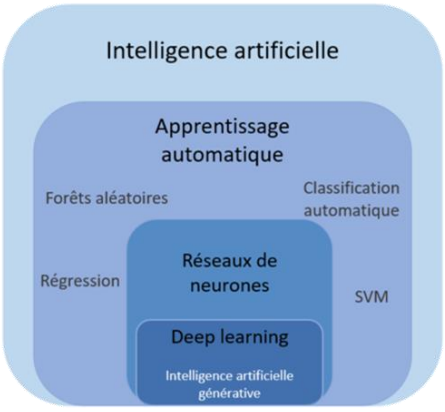
L'intelligence artificielle englobe un large spectre de technologies visant à créer des systèmes capables de réaliser des tâches nécessitant habituellement des capacités cognitives humaines. Le développement de ces technologies repose sur plusieurs disciplines scientifiques dont les statistiques, l'analyse de données, l'informatique et les sciences cognitives. L'intelligence artificielle est définie par l'*AI Act* de l'Union européenne[126] comme un système automatisé conçu pour fonctionner à différents niveaux d'autonomie, qui peut faire preuve d'une capacité d'adaptation après son déploiement et qui déduit, à partir des données d'entrée qu'il reçoit, la manière de générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels. Les débuts de l'intelligence artificielle remontent aux années 1940-1960[127] avec la création d'un premier modèle mathématique de réseau de neurones en 1943[128] et la publication d'un article par Alan Turing en 1950 s'interrogeant pour la première fois sur une intelligence potentielle des machines, dans lequel il imagine le test de Turing[129]. Le développement de l'intelligence artificielle a évolué avec les développements technologiques au cours des décennies suivantes et a connu une accélération marquée dans le domaine de l'apprentissage automatique (*machine learning*), avec les technologies d'apprentissage profond (*deep learning*)[130]. L'augmentation de la puissance de calcul des ordinateurs et l'accélération des travaux sur le sujet ont permis de développer les techniques d'intelligence artificielle avec des technologies de plus en plus sophistiquées et des applications de plus en plus nombreuses.

L'intelligence artificielle générative est, aujourd'hui, la branche de l'intelligence artificielle la plus connue du grand public. Celle-ci utilise des algorithmes d'apprentissage automatique basés sur le *deep learning* pour créer de nouveaux contenus sous différentes formes (texte, image, vidéo, son). En novembre 2022, la société OpenAI a lancé son prototype d'agent conversationnel, ChatGPT, utilisant l'intelligence artificielle générative pour créer du texte à partir de grands modèles de langage entraînés sur de larges volumes de texte (appelés, en anglais, *large language models*). ChatGPT utilise, pour son apprentissage, des données disponibles en ligne ainsi que des données fournies par l'utilisateur lors de ses interactions avec l'outil. Son lancement a suscité un fort engouement

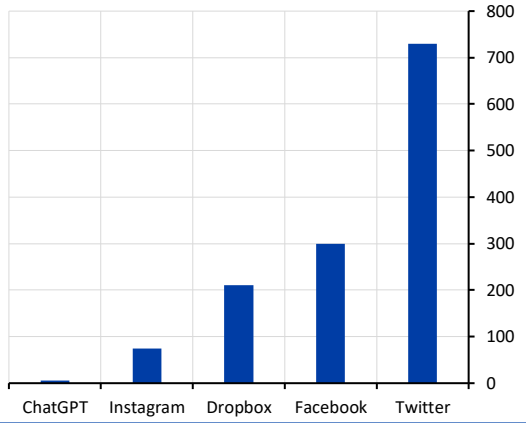
[125] Financial Stability Review - ECB May 2024
[126] AI Act | Shaping Europe's digital future (europa.eu)
[127] Histoire de l'intelligence artificielle - Intelligence artificielle (coe.int)
[128] MCCULLOCH, Warren S. PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 1943, vol. 5, p. 115-133.
[129] Le test de Turing est un test de l'aptitude d'une machine à imiter une conversation humaine de manière indistinguable d'un humain. TURING, Alan M. *Computing machinery and intelligence*.
[130] L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle, qui se concentre sur le développement de programmes informatiques conçus pour apprendre au fur et à mesure de leur développement et de leur utilisation. L'apprentissage profond est une branche de l'apprentissage automatique qui s'appuie sur la technique d'apprentissage des réseaux neuronaux.

du grand public et l'application a atteint 1 million d'utilisateurs uniques en seulement cinq jours, soit l'adoption la plus rapide de l'histoire des applications web.

Graphique 3.1 : Champs d'étude de l'intelligence artificielle



Graphique 3.2 : Nombre de jours pour qu'un service atteigne 1million d'utilisateurs depuis leur mise en ligne



Source : [Gauche] : Banque de France ; [Droite] : SimilarWeb;
Note : Mise en ligne de ChatGPT le 30 novembre 2022

L'intelligence artificielle a déjà de multiples applications dans de nombreux secteurs économiques

En 2022, 50 % des entreprises internationales indiquaient avoir déjà adopté les technologies basées sur l'intelligence artificielle (générative ou non) dans leurs activités (cf. graphique 3.3)**.** Ce sondage[131], réalisé auprès de plus de 1 400 participants, met en évidence que la part des entreprises utilisant l'intelligence artificielle aurait doublé entre 2017 et 2018 mais serait relativement stable au cours des dernières années. Cette étude suggère que le développement récent de l'intelligence artificielle serait principalement porté par une intensification des applications de la part des entreprises ayant déjà adopté l'intelligence artificielle. En France, selon une enquête récente[132], 35 % des entreprises déclarent utiliser des technologies liées à l'intelligence artificielle[133].

Ces technologies sont principalement utilisées afin d'optimiser les opérations de services, le développement de nouveaux produits et la gestion financière. L'intelligence artificielle est utilisée dans le secteur manufacturier notamment pour la gestion prédictive des stocks, la prévision de maintenance des équipements industriels, l'inspection qualité des produits et la gestion des déchets. L'intelligence artificielle bénéficie aussi au secteur agricole, grâce à des prévisions météorologiques affinées permettant d'anticiper des évènements météorologiques qui peuvent avoir un impact sur les récoltes. Dans le secteur des services, l'intelligence artificielle permet par exemple l'amélioration du service client (recommandations personnalisées), la tarification dynamique (tourisme et transports) et l'amélioration de la cybersécurité (surveillance des systèmes, analyse des vulnérabilités et des menaces en temps réel). Enfin, l'intelligence artificielle est aussi utilisée dans le domaine de la santé, pour l'aide à la décision et la chirurgie assistée par ordinateur.

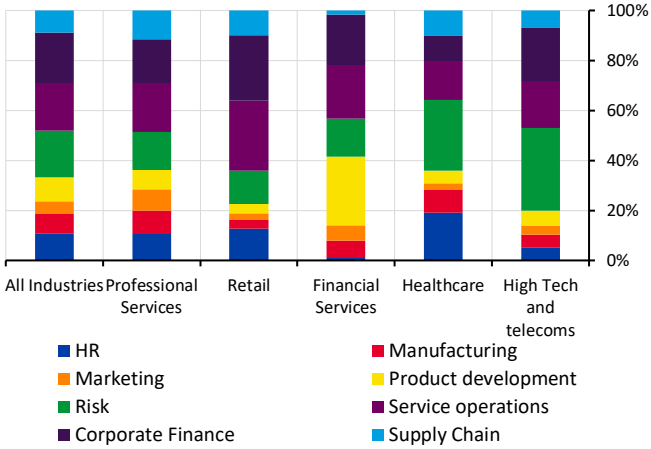
L'intelligence artificielle générative marque une accélération technologique qui amplifie les applications de l'intelligence artificielle. L'intelligence générative permet la génération simple et rapide de contenu et pourrait faciliter l'automatisation de certaines tâches, par exemple dans le domaine du service client, avec son intégration à des agents conversationnels (*chatbot*). L'intelligence artificielle générative a déjà suscité un intérêt marqué des entreprises (cf. graphique 3.4). À ce stade, la majorité des entreprises utilisant l'intelligence artificielle générative indiquent l'avoir déployée de façon limitée, tout en anticipant une augmentation de ses applications dans les prochaines années[134]. Parmi les principaux freins au déploiement de l'intelligence artificielle générative, les

[131] The state of AI in 2022—and a half decade in review | McKinsey
[132] Enquête « Les employeurs face à l'Intelligence Artificielle » auprès d'un échantillon de 3 000 établissements français de 10 salariés ou plus, juin 2023 – France Travail.
[133] Diagnostic structurel de la France 2023 (note d23-124)
[134] Generative AI: Differentiating disruptors from the disrupted | MIT Technology Review

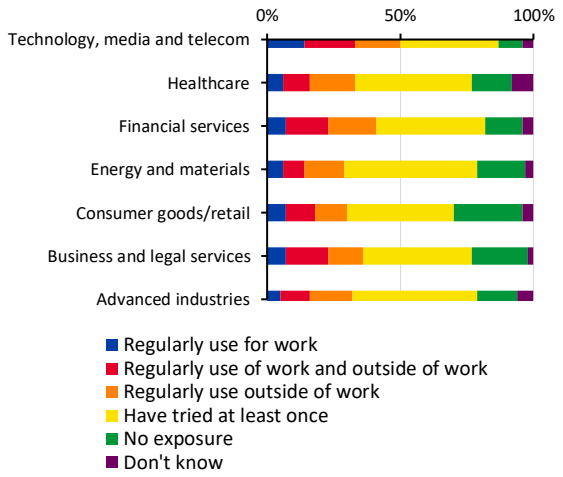
entreprises citent les enjeux de confidentialité des données, de confiance, de transparence et le manque de compétences[135].

L'intelligence artificielle pourrait avoir des impacts économiques significatifs, mais qui restent difficiles à prévoir. En permettant l'optimisation de la production de biens et de services, l'intelligence artificielle pourrait avoir un effet positif transitoire sur le taux de croissance de la productivité. D'après une étude menée en 2023 au sein d'un service clients aux États-Unis, l'utilisation de l'intelligence artificielle générative aurait permis une augmentation de l'ordre de 25 % en moyenne de la productivité des salariés cinq mois après l'introduction de la technologie, en bénéficiant le plus aux salariés les moins expérimentés ou les moins qualifiés[136]. De plus, l'intelligence artificielle pourrait également avoir un effet positif permanent sur le taux de croissance de la productivité en automatisant la production d'idées. Ces effets macroéconomiques restent difficiles à mesurer et leur ampleur dépendra des politiques mises en place[137]. Selon un rapport coordonné par Anne Bouverot et Philippe Aghion publié en mars 2024, en France, les gains de productivité générés par l'intelligence pourraient augmenter le PIB de 250 à 420 milliards d'euros[138]. Un article publié en décembre 2023 par le FMI[139] estime que les gains de croissance resteraient limités dans un processus d'automatisation qui viendrait remplacer le travail humain, mais qu'ils augmenteraient si le développement de l'intelligence artificielle se faisait de façon complémentaire.

Graphique 3.3 : Utilisation de l'intelligence artificielle par les entreprises internationales en 2022



Graphique 3.4 : Exposition aux outils d'intelligence artificielle générative dans les entreprises internationales



Sources : [Gauche] : Données Survey McKinsey 2022 ; [Droite] : Données McKinsey 2023

Le secteur financier a déjà largement adopté l'intelligence artificielle et devrait encore accroître son recours à ces technologies au cours des prochaines années

Depuis de nombreuses années, le secteur financier a fortement adopté l'intelligence artificielle pour ses activités. L'utilisation de l'intelligence artificielle dans les services financiers remonte aux années 1980, avec le lancement d'outils basés sur l'intelligence artificielle pour les conseils financiers et fiscaux personnalisés. Depuis, les institutions financières ont peu à peu intégré l'intelligence artificielle dans leurs activités de marché (prévisions, analyses financières et prises de décisions).

[135] Data Suggests Growth in Enterprise Adoption of AI is Due to Widespread Deployment by Early Adopters (ibm.com)
[136] Brynjolfsson E., Li D., Raymond L (2023), "Generative AI at Work", NBER working paper series.
[137] Trésor-Éco n° 341 (Avril 2024), " Les enjeux économiques de l'intelligence artificielle " (economie.gouv.fr)
[138] Commission de l'intelligence artificielle, 25 recommandations pour l'IA en France, rapport coordonné par Anne Bouverot et Philippe Aghion, 13 mars 2024.
[139] Acemoglu (D.) and Johnson (S.), Rebalancing AI, décembre 2023.

Les banques et assurances mobilisent déjà largement l'intelligence artificielle pour leurs activités. Schématiquement, quatre grandes familles de cas d'usage – et de bénéfices escomptés – peuvent être identifiées[140] :

- L'amélioration de l'expérience et de la satisfaction du client, au travers de divers systèmes d'assistance, d'aide à la décision ou de recommandation qui peuvent en particulier prendre la forme de *chatbots* ou de *voicebots*. Les capacités de personnalisation de l'intelligence artificielle, encore accrues par l'intelligence artificielle générative, permettent notamment d'améliorer l'efficacité des processus de marketing et de vente.
- L'optimisation de la chaîne de valeur : optimisation des processus internes et réduction des coûts, grâce à l'automatisation de tâches auparavant manuelles. Les effets produits sont d'autant plus forts que l'intelligence artificielle offre la possibilité de traiter des données non structurées. Le cas emblématique dans le domaine bancaire est l'essor de l'autonomisation des clients pour la gestion courante, ce qui permet progressivement aux conseillers en agence de se départir des tâches à moindre valeur ajoutée.
- La gestion des risques, qu'il s'agisse des risques financiers (risque de crédit, par exemple), des risques opérationnels ou des risques de conformité. La lutte contre la fraude et la lutte contre le blanchiment des capitaux et le financement du terrorisme (LCB-FT) constituent des domaines d'application privilégiés, par des systèmes de suivi des transactions suspectes et de gestion automatisée des alertes.
- Les activités de marché constituent également un domaine privilégié d'application des technologies d'intelligence artificielle, pour permettre l'amélioration de l'identification de tendances et intégrer d'importants volumes de données. Les participants utilisent ainsi l'apprentissage automatique dans le cadre du *trading* algorithmique, par exemple pour identifier la meilleur stratégie d'exécution d'une transaction en fonction des conditions de marché. Ces technologies peuvent aussi être utilisées par les gestionnaires d'actifs, pour la construction de portefeuilles ou l'évaluation des actifs. Néanmoins, certaines institutions financières font des déclarations erronées auprès de leurs clients concernant leur utilisation de l'intelligence artificielle et deux entreprises financières ont été sanctionnées par la Securities Exchange Commission pour ces pratiques[141].

Dans la pratique, de nombreux cas d'usage combinent tout ou partie de ces caractéristiques. Ainsi, dans le domaine bancaire, les systèmes basés sur l'apprentissage automatique permettent par exemple de déterminer la solvabilité des clients à partir de l'étude de leurs données financières (en produisant des cotes de crédit telles que le *credit score* aux États-Unis). Cela conduit à réduire les délais de traitement des demandes de prêts (amélioration de l'expérience client, meilleure efficacité des processus de distribution), par des procédés largement automatiques (optimisation des processus internes), tout en améliorant l'anticipation des défauts de paiement par une meilleure modélisation (réduction des risques)[142]. De même, en assurance, les systèmes d'intelligence artificielle sont utilisés pour accroître la personnalisation de la tarification suivant le profil du client, dans la limite de la règlementation ou encore afin de réduire les délais et les coûts de traitement des sinistres, par des analyses automatiques d'images ou des fonctions de lecture avancée des rapports d'expertise. Un sondage mené par Moody's Analytics en novembre 2023 a mis en évidence qu'à ce stade 30 % des institutions financières interrogées sont en phase d'essai ou utilisent activement l'intelligence artificielle pour leurs activités de gestion des risques et de conformité[143].

L'essor de l'intelligence artificielle générative devrait trouver de nombreuses applications dans le secteur financier. Si l'intelligence artificielle est déjà largement utilisée par les institutions financières, le déploiement de

[140] Voir notamment les deux études de l'Autorité de contrôle prudentiel et de résolution (ACPR) sur la transformation numérique du secteur financier français : La transformation numérique dans le secteur bancaire français, Analyses et Synthèses n°131, janvier 2022 ; La transformation numérique dans le secteur français de l'assurance, Analyses et Synthèses n°132, janvier 2022.

[141] SEC.gov | SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence Delphia et Global Predictions ont fait croire à leurs clients et clients potentiels qu'ils utilisaient l'IA, alors que ce n'était pas le cas. La SEC a constaté que ces conseillers en investissement ont prétendu utiliser un modèle d'IA alors que ce n'était pas le cas et ont induit leurs clients en erreur.

[142] Artificial intelligence and machine learning in financial services - FSB 2017

[143] AI in compliance (moodys.com)

l'intelligence artificielle générative dans le secteur financier semble limité et principalement orienté vers l'amélioration des processus internes (identification des risques, développement de code, génération et synthèse de documents, lutte contre la fraude et le blanchissement d'argent). Selon une enquête de décembre 2023[144], une large majorité d'institutions financières anticipe une hausse des applications liée à l'intelligence artificielle générative sous trois ans avec des impacts significatifs pour leur secteur d'activité. Un développement plus large pourrait concerner les domaines en contact direct avec les clients[145]. Par exemple, contrairement à l'intelligence artificielle intégrée à ce stade dans les *chatbots* permettant uniquement des réponses prédéfinies, l'intelligence artificielle générative pourra fournir des informations personnalisées à l'utilisateur comme son solde bancaire, ses derniers virements effectués ou encore lui proposer des produits d'épargne personnalisés. Les *robo-advisors* (« robots-conseillers »), plateformes en ligne qui livrent des conseils financiers et permettent parfois la gestion automatisée d'un portefeuille d'actifs constituent un exemple d'application de l'intelligence artificielle « classique », qui pourrait être complétée par l'intelligence artificielle générative afin d'améliorer le recueil des informations client, de proposer des produits et des conseils personnalisés en language naturel[146]. Le développement de ces applications destinées aux clients nécessite de prendre en compte les enjeux de transparence et de responsabilité.

Les technologies basées sur l'intelligence artificielle peuvent faciliter les missions de supervision du système financier (démarche dite « SupTech », pour *supervisory technologies*). L'intelligence artificielle offre des opportunités d'amélioration de l'efficacité des superviseurs dans l'exercice de leurs missions, dont des solutions de contrôle des risques[147]. La Banque de France et l'Autorité de contrôle prudentiel et de résolution (ACPR) ont ainsi développé plusieurs projets en ce sens (cf. encadré). De même, l'Autorité monétaire de Singapour (MAS) utilise l'intelligence artificielle pour analyser des documents afin d'identifier des points d'attention et pour l'analyse des données de transactions afin de détecter les potentielles manipulations de marché[148] et la *Securities and Exchange Commission* (SEC) utilise l'intelligence artificielle pour repérer des activités frauduleuses et des manipulations de marché.

3.2 Des risques inhérents à la nature de l'intelligence artificielle

Si les potentielles opportunités de l'intelligence artificielle sont nombreuses, son déploiement comporte des risques inhérents à sa nature dont certains pourraient être renforcés par l'intelligence artificielle générative. En premier lieu, l'intelligence artificielle soulève des enjeux liés à la qualité et à la confidentialité des données utilisées pour son apprentissage. **Le fonctionnement autonome et opaque des modèles d'intelligence implique également le risque d'une compréhension difficile des résultats fournis et de leur fiabilité. Enfin, une concentration excessive des acteurs du secteur exacerberait ces risques pour les utilisateurs et serait préjudiciable au système dans son ensemble.**

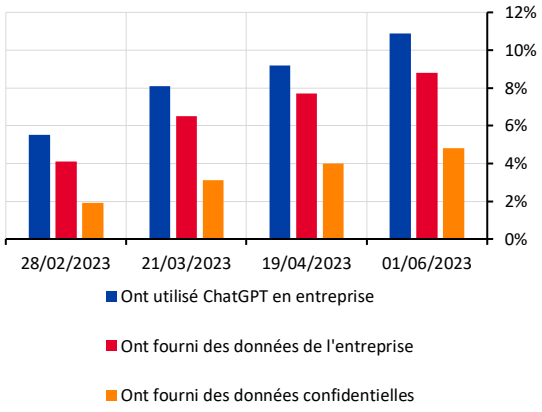
L'intelligence artificielle, par le fonctionnement de ses algorithmes et son apprentissage, engendre des risques liés à la gestion des données notamment leur confidentialité. Ce risque est particulièrement marqué dans les systèmes d'intelligence artificielle générative qui apprennent, non seulement, à partir des données disponibles sur Internet, mais aussi à partir des données fournies par les utilisateurs (en vue d'améliorer leurs réponses) : ces outils peuvent ainsi transmettre les informations apprises d'un utilisateur en générant des contenus pour d'autres utilisateurs. D'après un rapport de l'entreprise spécialisée Cyberhaven publié en juin 2023[149], 10,8 % des employés ont utilisé ChatGPT au travail et 4,7 % d'entre eux ont déjà fourni des informations confidentielles de leur entreprise à l'agent conversationnel, le pourcentage ayant même tendance à croître au cours du temps. En juin 2023, les données confidentielles représentaient 11 % du total des données copiées et collées directement dans l'outil. Ces comportements représentent un danger pour les entreprises, dont les informations sensibles peuvent

[144] 2023 IIF-EY Survey Report on AI ML Use in Financial Services - Public Report - Final.pdf
[145] GENERATIVE ARTIFICIAL INTELLIGENCE IN FINANCE - December 2023 (oecd-ilibrary.org)
[146] Les *robo-advisors* fonctionnent avec des systèmes de règles, proposant une allocation d'actifs à partir des informations transmises par le client ou avec des algorithmes complexes permettant la surveillance continue du portefeuille et son ajustement dynamique. Voir notamment : Robo-Advising : Moins d'IA et plus de XAI ? - 2021 (institutlouisbachelier.org)
[147] Le superviseur financier à l'heure de l'IA | Banque de France (banque-france.fr)
[148] Written reply to Parliamentary Question on use of artificial intelligence in supervision of financial institutions (mas.gov.sg)
[149] Rapport créé à partir de données provenant d'1,6 million d'utilisateurs des logiciels de Cyberhaven, au niveau mondial.

être diffusées largement. En outre, les données transmises à des outils conversationnels peuvent aussi être volées lors de cyberattaques. Pour pallier ce risque et préserver la confidentialité des données financières, plusieurs groupes bancaires dont JPMorgan, Bank of America, Goldman Sachs et Deutsche Bank ont interdit l'utilisation du service en ligne ChatGPT à leurs employés. Néanmoins, ces risques pourraient être moindre dans le cas d'outils réservés à l'usage interne dans les entreprises, avec des règles strictes de gestion des données.

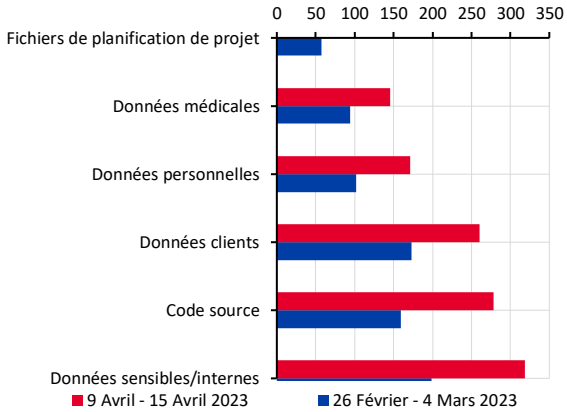
Graphique 3.5 : Adoption de ChatGPT par les salariés

x : axe temporel / y : pourcentage cumulatif



Graphique 3.6 : Nombre d'incidents de transmission de données sensibles à ChatGPT

x : nombre d'incidents par 100 000 employés / y : type d'incident



Source : Cyberhaven

Plus généralement, la qualité des résultats obtenus par des technologies d'intelligence artificielle dépend directement de la qualité des données. Le recours à des données inadéquates lors de l'apprentissage de l'outil, ou lors de son utilisation, peut entraîner des résultats erronés[150]. Certaines technologies sont limitées par leur historique de données ou bien par les sources de données disponibles. À titre d'exemple, GPT 3.5 avait été entraîné sur l'historique des données d'Internet jusqu'à janvier 2022, un utilisateur peut donc obtenir des informations erronées ou obsolètes en interrogeant l'outil (pas de prise en compte de la guerre en Ukraine, par exemple). À ce stade, l'information générée par les outils d'intelligence artificielle générative provient, pour l'essentiel, de contenus rédigés par des humains. Or, une forte adoption de ces outils pourrait mener à un phénomène de boucle : le contenu qu'ils génèrent pourrait devenir l'une des principales sources d'information des modèles d'intelligence artificielle générative. Ce phénomène de récursivité pose un risque de perte de diversité de l'information originale au profit d'une standardisation des réponses fournies potentiellement biaisées, erronées ou obsolètes[151].

L'intelligence artificielle peut perpétuer les biais historiques provenant des données ou en créer de nouveaux en raison de son fonctionnement. Ceci peut créer des points aveugles, voire entraîner un manque d'équité. Un modèle entraîné sur des données biaisées perpétuera – voire renforcera – ces biais historiques et potentiellement des discriminations. Et ce, d'autant plus qu'un modèle biaisé génère des résultats biaisés qui seront ensuite utilisés dans son processus d'apprentissage et renforceront les biais (par récursivité). À titre d'exemple, l'algorithme utilisé par Apple et Goldman Sachs, pour les décisions d'allocation de lignes de crédit, aurait offert des prêts de montants inférieurs aux femmes qu'aux hommes, toutes choses égales par ailleurs[152]. En outre, les décisions biaisées prises par les modèles d'apprentissage automatique peuvent se produire même avec des données de bonne qualité, en raison d'approximations, de défauts dans la construction de l'algorithme, ou du fait que les corrélations entre variables peuvent être spécifiques aux bases de données utilisées (reflétant, par exemple, les caractéristiques d'un échantillon historique).

En outre, l'intelligence artificielle comporte des risques liés aux modèles qui peuvent remettre en cause la fiabilité des résultats. Par exemple, les « hallucinations » sont des réponses fausses présentées comme des faits

[150] Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers (oecd.org)
[151] The curse of recursion: training on generated data makes models forget. SHUMAILOV, Ilia, SHUMAYLOV, Zakhar, ZHAO, Yiren, et al. 2023.
[152] Bias, fairness, and other ethical dimensions in artificial intelligence – Bank Underground

par des outils utilisant l'intelligence artificielle générative avec des grands modèles de langage. Ces problèmes d'hallucinations de modèles d'intelligence artificielle générative sont inhérents à la nature des grands modèles de langage, qui estiment une réponse probable à partir des mots fournis par l'utilisateur et peuvent donner des réponses anormales, voire fausses. À titre d'exemple, une entreprise a demandé à ChatGPT de produire un article à propos des résultats de l'entreprise Tesma, ChatGPT a créé un article cohérent mais avec des chiffres financiers inventés et complètement faux. Ce phénomène peut aussi se produire lorsque le modèle est entraîné sur des données fausses. Par ailleurs, une même question posée à deux moments différents peut générer des réponses différentes, mettant en évidence un problème de cohérence temporelle. Un article de la Banque des règlements internationaux (BRI), publié en janvier 2024, illustre également les limites des modèles d'intelligence artificielle générative, qui inventent des réponses fausses lorsqu'ils font face à des problèmes nouveaux[153].

De fait, les technologies utilisant l'intelligence artificielle produisent des résultats complexes et leur fonctionnement opaque implique des difficultés de compréhension des résultats. Les résultats des systèmes basés sur l'intelligence artificielle sont intrinsèquement complexes à interpréter pour un esprit humain, en raison de la nature de certaines modélisations utilisées (réseaux de neurones, etc.), qui reposent essentiellement sur des relations non linéaires. Les acteurs du secteur de l'intelligence artificielle peuvent parfois eux-mêmes renforcer l'opacité du fonctionnement des outils, afin de protéger leur propriété intellectuelle. Certains outils d'intelligence artificielle constituent ainsi des boîtes noires, dont les paramètres et le fonctionnement ne sont pas visibles pour les entités les intégrant dans leurs services et dont les sources ne sont pas mentionnées.

Une gouvernance claire est nécessaire pour réduire les risques associés à de potentiels biais, au manque de transparence et à la fiabilité de l'outil. Une gouvernance claire nécessite de définir des lignes de responsabilité pour le développement et la supervision des systèmes basés sur l'intelligence artificielle tout au long de leur cycle de vie, du développement au déploiement, et la désignation explicite de la responsabilité des résultats produits par le modèle.

Le manque de concurrence entre les fournisseurs d'outils basés sur l'intelligence artificielle peut exacerber les risques présentés. Les coûts nécessaires au développement de modèles basés sur l'intelligence artificielle peuvent impliquer un risque de dépendance à des fournisseurs oligopolistiques. À ce stade, OpenAI représente 39 % du marché de l'intelligence artificielle générative contre 30 % pour Microsoft, 8 % pour Amazon et 7 % pour Google[154]. La concentration de ces technologies chez un nombre limité de fournisseurs (s'appuyant sur leurs avantages en termes de données ou de puissance de calcul) pourrait amplifier le risque de récursivité d'utilisation de données fournies par des intelligences artificielles pour l'apprentissage des modèles. Les modèles seraient entraînés par des données provenant d'un nombre limité d'outils basés sur l'intelligence artificielle et perdraient en diversité.

L'intelligence artificielle générative peut renforcer les risques de cyberattaques. Même si la plupart des agents conversationnels sont créés avec des mesures de sécurité internes, comme le refus de répondre à une demande directe de création de cyberattaque, ils ne sont généralement pas programmés pour refuser la création de code informatique ou de texte pouvant indirectement servir à des cyberattaques.

3.3 L'intelligence artificielle présente des risques pour la stabilité financière

Ces risques inhérents à l'intelligence artificielle constituent des facteurs potentiels d'amplification de risques existants pour la stabilité financière. En premier lieu, l'utilisation des modèles similaires d'intelligence artificielle par une large partie des participants de marché pourrait renforcer les risques de procyclicité et de volatilité des marchés. Par ailleurs, le déploiement de l'intelligence artificielle pourrait entraîner une concentration des marchés et favoriser l'émergence de nouveaux acteurs d'importance systémique. Enfin, elle renforce le risque de cyberattaques auquel le système financier est particulièrement exposé.

[153] Testing the cognitive limits of large language models (bis.org)
[154] The leading generative AI companies (iot-analytics.com)

L'intelligence artificielle pourrait accentuer la procyclicité et la volatilité des marchés. Le développement du recours à l'intelligence artificielle pour des stratégies de négociation ou des décisions d'investissement peut créer un risque d'homogénéité des recommandations, ainsi que des comportements moutonniers pouvant mener à un risque accru de procyclicité pour les marchés[155]. L'utilisation des mêmes outils d'intelligence artificielle par de nombreux acteurs de marché peut engendrer un cumul de transactions unidirectionnelles, et donc amplifier les mouvements de marché en cas de mouvement adverse. Si un algorithme recommande à plusieurs acteurs significatifs de vendre un actif, alors la baisse de prix consécutive peut engendrer des recommandations de vente pour d'autres acteurs, et ainsi alimenter la chute du cours de l'actif avec des impacts sur les opérations de couverture associées, et ce, de manière extrêmement rapide (*flash crash*). Ces épisodes de tensions peuvent entraîner d'importants besoins de liquidité pour les acteurs les plus fragiles, qui peuvent notamment se traduire par des ventes forcées d'actifs, avec pour conséquence potentielle une extension des tensions sur d'autres classes d'actifs. Compte tenu des fortes interconnexions directes ou indirectes des institutions financières par le biais des marchés financiers, de tels mouvements peuvent donc avoir des effets systémiques. De plus, en cas de manque de compréhension des modèles, les utilisateurs pourraient avoir des difficultés à prédire leur réaction en fonction des conditions de marché et à ajuster leur stratégie en période de faible performance. En outre, l'intelligence artificielle pourrait renforcer le risque de manipulation de marché. Par exemple, la diffusion de fausses informations facilitée par les technologies d'intelligence artificielle générative permettant de générer des images et du son à partir de contenu existant (appelés *deepfakes*) pourrait engendrer des mouvements de marché massifs et une forte volatilité des marchés.

Plus généralement, une mauvaise estimation des risques pourrait engendrer d'importantes pertes financières pour les institutions financières. Une mauvaise estimation pourrait se traduire par l'octroi excessif de prêts, ou par des erreurs massives de tarification, conduisant à d'importantes pertes pour les institutions financières. De même que pour les modèles statistiques classiques, l'entraînement des modèles d'intelligence artificielle sur des données par définition non exhaustives peut conduire à une amplification des pertes en période de crise. En effet, les modèles étant généralement entraînés sur des données historiques, celles-ci ne sont pas nécessairement représentatives du comportement d'un actif donné lors d'évènements extrêmes. De même, l'utilisation d'un modèle avec des données non exhaustives peut engendrer une mauvaise tarification des prêts par les banques ou des pertes systématiques sur des produits d'assurance. De telles erreurs pourraient potentiellement se traduire par une perte de confiance des clients, avec des mouvements de type panique bancaire, et pourraient se présenter à une échelle systémique en cas d'utilisation de modèles similaires par les institutions financières.

Le déploiement de l'intelligence artificielle pourrait se traduire par une concentration des fournisseurs de services, avec des risques de dépendance, et par une concentration du système financier. À l'instar d'une utilisation massive du même logiciel par les acteurs de marché, l'utilisation généralisée d'un petit nombre d'outils d'intelligence artificielle peut engendrer un risque opérationnel en cas de dysfonctionnement ou d'indisponibilité de la technologie[156] les empêchant de mener leurs activités. De plus, la prépondérance d'un fournisseur pourrait amplifier les interconnexions entre les acteurs du système financier et la diffusion de chocs[157]. La domination croissante d'une poignée de systèmes de gestion des risques basés sur l'intelligence artificielle, comme *Aladdin* de BlackRock et *RiskMetrics* de MSCI, peut renforcer cette tendance par une augmentation de la concentration des acteurs spécialisés[158]. En outre, le développement d'outils basés sur l'intelligence artificielle nécessite des investissements initiaux importants et les grandes institutions financières sont plus à même de développer ces technologies que les petits acteurs. Ainsi, ces différences technologiques pourraient conduire à une plus grande concentration au sein du système financier.

L'intelligence artificielle générative pourrait renforcer le risque cyber, auquel le système financier est très exposé. En effet, l'intelligence artificielle générative permet aux cyberattaquants de créer des attaques plus sophistiquées et plus difficiles à détecter. Cette technologie permet, par exemple, la rédaction automatisée de

[155] Intelligence artificielle : enjeux pour le secteur financier - Olivier FLICHE, Su YANG - Pôle Fintech-Innovation, ACPR
[156] Humans keeping AI in check – emerging regulatory expectations in the financial sector (bis.org)
[157] SEC.gov | "AI, Finance, Movies, and the Law" Prepared Remarks before the Yale Law School
[158] Artificial intelligence and systemic risk (lse.ac.uk)

mails d'hameçonnage (*phishing*) personnalisés et à grande échelle dans plusieurs langues et sans faute d'orthographe. De même, les systèmes d'intelligence artificielle générative permettent d'imiter la voix (voire l'image) d'une personne en vue d'usurper son identité. L'intelligence artificielle générative facilite aussi les attaques reposant sur du code informatique malveillant. Les agents conversationnels sont en effet capables de produire ou de corriger du code informatique, ce qui permet de réaliser des attaques plus sophistiquées. L'intelligence artificielle permet aussi l'industrialisation des cyberattaques en automatisant leur conception et leur exécution car celle-ci est capable d'adapter sa stratégie en fonction des vulnérabilités détectées. En outre, l'intelligence artificielle générative permet une nouvelle catégorie de cyberattaque, nommée *indirect prompt injection,* qui consiste à manipuler les requêtes pour contourner la sécurité intégrée de l'outil afin que l'intelligence artificielle générative récupère elle-même des informations sensibles. Ces attaques sont possibles dès lors qu'une intelligence artificielle a la capacité de se connecter sur des serveurs par une requête, dont elle peut se servir pour transmettre des informations issues de sa conversation avec l'utilisateur. À la différence des méthodes précédentes, l'outil d'intelligence artificielle n'aide pas à produire des outils malveillants mais elle se comporte directement en acteur malveillant.

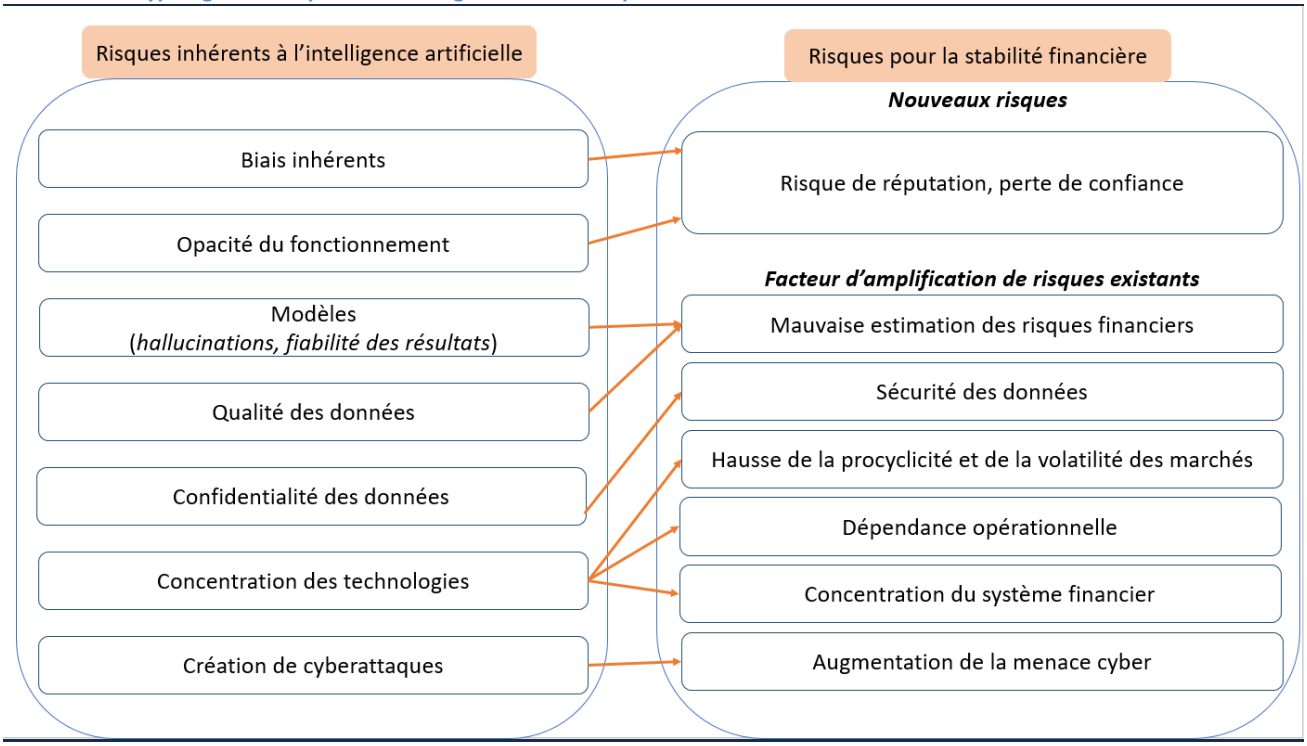
Néanmoins, des solutions d'intelligence artificielle générative sont en cours de développement pour assister dans la gestion du risque cyber, pour la détection des menaces et le renforcement dynamique des mesures de cybersécurité. La préparation face aux cyberattaques est un élément essentiel de la résilience cyber, et l'intelligence artificielle permet notamment d'analyser les vulnérabilités en direct, de détecter des cyberattaques, de simuler des scénarios d'attaques. L'intelligence artificielle offre aussi la possibilité d'une gestion automatisée des attaques, avec des protocoles de résolution adaptés à la menace.

Les risques inhérents à l'intelligence artificielle, notamment les risques liés aux biais potentiels, aux erreurs ou à l'opacité des modèles, pourraient également avoir des impacts réputationnels sur les institutions financières. Des erreurs ou des biais pourraient devenir systématiques à l'échelle d'un organisme ou d'un marché avec l'utilisation de modèles basés sur l'intelligence artificielle et engendrer une potentielle perte de confiance de la part des clients. Au-delà des risques pour les institutions financières, les biais potentiels et les difficultés de compréhension des résultats fournis peuvent engendrer un risque pour la protection des clients et des investisseurs. Une étude de l'ACPR met en évidence que les explications fournies sous forme de conversation par des *robo-advisors* tendent à augmenter, à tort, la confiance des clients (davantage prêts à suivre les recommandations incorrectes de l'outil) par rapport à des explications sous d'autres formes[159]. Ceux-ci peuvent donc être mal informés des risques liés à un investissement et prendre des décisions inadaptées à leur situation.

La capacité de compréhension des résultats est un facteur essentiel pour une bonne gestion des risques par les établissements financiers et pour les superviseurs. La compréhension du fonctionnement de l'algorithme est nécessaire pour vérifier non seulement les calculs effectués[160], mais aussi pour étudier les variables prises en compte et le poids assigné à chacune d'entre elles pour l'obtention du résultat[161]. En outre, la compréhension du fonctionnement est une condition nécessaire pour la vérification par les superviseurs que les calculs correspondent aux exigences réglementaires. Dans le cas des modèles internes d'évaluation de risque de crédit des banques, les algorithmes utilisés doivent être facilement interprétables et compréhensibles à la fois par la direction de la banque, afin qu'elle puisse prendre des décisions stratégiques en matière d'allocation de portefeuille de façon éclairée, et par le superviseur en charge de l'application des réglementations prudentielles[162].

[159] La motivation du conseil par les robo-advisors : vers un éclairage apporté aux clients ? | ACPR (banque-france.fr)
[160] Machine Learning et Modèles IRB : Avantages, Risques et Préconisations - Institut Louis Bachelier
[161] Artificial intelligence in central banking (bis.org)
[162] Artificial intelligence and machine learning in financial services - Financial Stability Board (fsb.org)

Schéma 3.1 : Typologie des risques de l'intelligence artificielle pour la stabilité financière



Source : Banque de France

3.4 Des risques structurels de manipulation de l'information pourraient avoir un impact sur le système financier

Les outils d'intelligence artificielle générative facilitent la création de fausses informations pouvant être utilisées à des fins de fraude et de déstabilisation géopolitique, avec des effets potentiellement déstabilisateurs pour le système financier.

Les technologies d'intelligence artificielle générative permettent de générer des images et du son à partir de contenu existant pour créer de fausses vidéos et enregistrements à des fins de manipulation et d'escroquerie. Ces contenus sont basés sur un processus appelé *deepfake* en anglais, qui consiste à superposer des images et du son pour générer un contenu falsifié. Début février 2024, l'image et la voix du gouverneur de la banque centrale de Roumanie ont été usurpées pour réaliser une vidéo visant à promouvoir des investissements frauduleux. En septembre 2023, le président de la *Securities and Exchange Commission* (SEC), Gary Gensler, a alerté sur le risque d'importants mouvements de marché unidirectionnels causés par les *deepfakes*, en cas de création de fausses informations menant à une perte de confiance dans un actif.

Ces fausses informations créées par les outils d'intelligence artificielle générative peuvent conduire à une perte de confiance vis-à-vis des institutions financières, et entraîner de potentielles paniques bancaires. La crise des banques régionales aux États-Unis, en mars 2023, a mis en exergue l'accélération de la panique des déposants par le biais de messages échangés sur les réseaux sociaux[163]. Dès lors que l'intelligence artificielle générative permet de rendre crédibles de fausses informations concernant des institutions bancaires, leur diffusion peut mener à une panique et à des retraits massifs des dépôts.

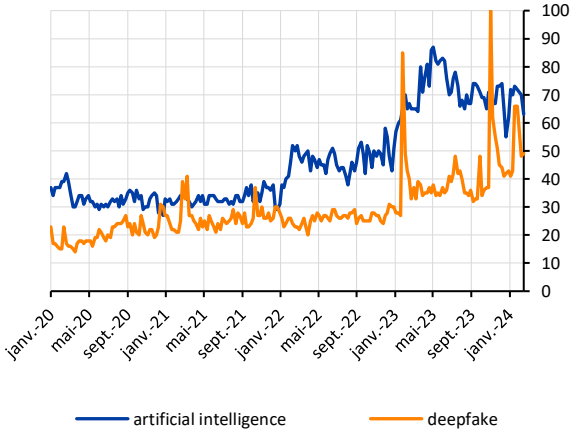
Plus généralement, les campagnes de désinformation peuvent entraîner une déstabilisation géopolitique pouvant engendrer d'importantes conséquences politiques, sociales et économiques. L'édition 2024 du rapport

[163] Social Media as a Bank Run Catalyst by J. Anthony Cookson, Corbin Fox, Javier Gil-Bazo, Juan Felipe Imbet, Christoph Schiller :: SSRN

sur les risques mondiaux du Forum économique mondial (FEM) classe le risque de désinformation en tête des menaces les plus sérieuses à court terme (deux ans)[164]. Pour limiter le risque de désinformation du grand public avec des *deepfakes* générés par intelligence artificielle générative, de plus en plus d'entreprises du secteur ont décidé de mettre en place des mesures pour faciliter la détection des *deepfakes* ou limiter leur diffusion. C'est le cas d'OpenAI qui a annoncé l'ajout de filigranes sur les images générées par ses systèmes, ou de MidJourney qui a interdit la génération d'images utilisant le visage de personnalités politiques. Pour améliorer la détection de *deepfakes* et empêcher leur diffusion, la sensibilisation du public est nécessaire, ainsi que la collaboration des différents acteurs (entreprises spécialisées, réseaux sociaux, opérateurs téléphoniques, médias).

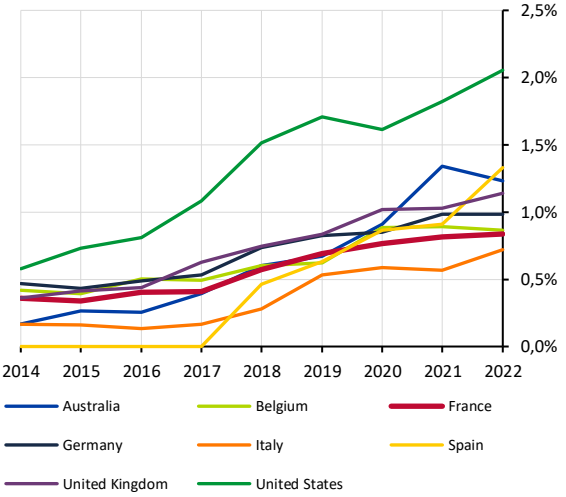
Graphique 3.7 : Volumes de recherches Google des termes *intelligence artificielle* et *deepfakes*

x : axe temporel / y : indice



Graphique 3.8 : Part des emplois en intelligence artificielle dans l'ensemble des offres d'emploi

x : axe temporel / y : pourcentage



Source : [Gauche] :Google Trends ; [Droite] : OurWorldInData
Note : Les tendances Google présentées ici sont mesurées sous la forme d'un indice qui prend la valeur 100 pour le point d'intérêt de recherche le plus élevé pour le terme depuis début 2020 dans le monde.

L'automatisation de certaines tâches, permise par l'intelligence artificielle, pourrait conduire à des changements structurels pour l'économie, même si ces effets restent difficiles à estimer avec certitude[165]. Tous les pays et toutes les fonctions économiques ne sont pas également exposées aux changements du marché du travail par l'intelligence artificielle. Les économies développées disposent, d'une part, d'une infrastructure informatique permettant le déploiement de l'intelligence artificielle et, d'autre part, de nombreuses fonctions économiques automatisées ou automatisables (robots dans le secteur industriel, importance des fonctions juridiques, comptables, financières et informatiques) ; ce sont donc les plus susceptibles d'être affectées par l'essor de l'intelligence artificielle[166]. Néanmoins, à court terme, peu de fonctions économiques seront entièrement automatisées par l'intelligence artificielle, et le plus probable est que nombre de métiers intègrent progressivement des composants de ces technologies.

L'intelligence artificielle est ainsi susceptible de conduire à des restructurations sectorielles majeures, qui pourraient alors déstabiliser le secteur financier si elles étaient mal anticipées ou plus rapides que prévu. Le déploiement de l'intelligence artificielle pourrait entraîner d'importants transferts d'emplois de certains secteurs économiques vers d'autres et, par conséquent, mener à un nombre élevé de faillites dans certains secteurs, ou au sein des entreprises elles-mêmes, avec des effets potentiels sur la solvabilité des entreprises ou des ménages concernés. Enfin, sur les marchés, l'adoption croissante et l'engouement autour de l'intelligence artificielle ont fait grimper les valorisations des entreprises liées au secteur en raison de l'optimisme des investisseurs quant aux bénéfices futurs de ces nouvelles technologies (cf. encadré 1.2 chapitre 1)[167].

[164] The world is changing and so are the challenges it faces | World Economic Forum (weforum.org)
[165] L'innovation par les banques centrales : le plus tôt est le mieux | Banque de France (banque-france.fr)
[166] Gen-AI: Artificial Intelligence and the Future of Work (imf.org)
[167] Financial Stability Review, November 2023 (europa.eu)

Le développement de l'intelligence artificielle dans le futur devra prendre en compte les limites imposées par les contraintes environnementales. L'entraînement d'un modèle d'intelligence artificielle nécessite une puissance de calcul importante, qui requiert à son tour une grande quantité d'énergie et implique des émissions de gaz à effet de serre. D'après les recherches d'OpenAI[168], depuis 2012, la puissance de calcul nécessaire pour l'apprentissage des modèles d'intelligence artificielle a doublé tous les trois mois. Une recherche en ligne utilisant l'intelligence artificielle générative utilise environ 4 à 5 fois l'énergie nécessaire à une recherche en ligne classique sur un moteur de recherche[169]. En outre, les systèmes d'intelligence artificielle générative nécessitent d'importantes quantités d'eau afin de refroidir les processeurs. Le développement des modèles d'intelligence artificielle générative par Google et Microsoft a engendré des augmentations annuelles de consommation d'eau de 20 % et 34 %, respectivement, d'après les rapports environnementaux des entreprises. Les conséquences environnementales de l'intelligence artificielle seront donc des limites potentielles à son déploiement.

Encadré 3.1 : Utilisation de l'intelligence artificielle par la Banque de France

Par Philippe Grad

L'essor de l'intelligence artificielle présente également des opportunités pour l'exercice des missions de la Banque de France. Afin de tirer le meilleur parti de ces nouvelles technologies, tout en en maîtrisant les risques, la Banque de France a adopté une démarche d'expérimentation. Celle-ci consiste à tester l'apport potentiel d'une solution technologique donnée sur un cas d'usage spécifique, afin de bien en appréhender les mérites et les inconvénients, et de lever au maximum les incertitudes. À l'issue de la phase d'expérimentation, si celle-ci se révèle concluante, le projet passe en production.

Dans ce contexte, une dizaine d'applications intègrent aujourd'hui des fonctionnalités d'intelligence artificielle, parmi lesquelles :

- **N-ACSEL :** modèle de partitionnement (*clustering*) basé sur des données économiques qui permet de « profiler » les territoires. Cette vision, restituée aux responsables locaux, leur permet de mieux connaître leurs forces et leurs faiblesses et de se comparer aux autres territoires ;
- **ScoreIA :** modèle de cote de crédit, incorporé dans l'application d'aide à la cotation des entreprises afin d'être utilisé comme un outil d'aide à la décision par les analystes ;
- **PAI :** outil de mesure des perceptions et des anticipations d'inflation, établi par analyse d'un large panel d'articles de presse ;
- **BASTID :** système de détection de fraudes bancaires ;
- **LUCIA :** un outil pour la lutte contre le blanchiment (LCB-FT), permettant notamment l'analyse de grands volumes d'opérations bancaires et utilisé dans le cadre des missions de contrôle sur place des établissements ;
- **PLUME :** système de retranscription d'enregistrements vocaux vers du texte, utilisé notamment dans le contrôle des pratiques commerciales du secteur financier.

D'autres travaux, plus exploratoires, utilisant de l'intelligence artificielle et d'autres sources originales comme les données satellites, par exemple, sont aussi menés sur l'utilisation des données de pollution pour suivre l'activité industrielle[170] y compris avec des partenaires extérieurs[171], permettant de couvrir de nombreux pays avec des méthodes homogènes.

[168] AI and Compute (georgetown.edu)
[169] Generative AI's environmental costs are soaring — and mostly secret (nature.com)
[170] Can satellite data on air pollution? - Banque de France - Novembre 2021
[171] Satellites Turn "Concrete": Tracking Cement with Satellite Data and Neural Networks - Banque de France - June 2023

La Banque de France bénéficie en effet d'un contexte favorable pour déployer les solutions à base d'intelligence artificielle : des données variées et nombreuses, une plateforme pour centraliser ces données et organiser leur gouvernance, des infrastructures pour l'innovation mais aussi une large expertise technique, notamment dans les équipes de *data scientists* et *data engineers* rattachées aux différents métiers, qui travaillent en coordination avec le LAB d'innovation de la Banque de France.

En parallèle de ces réalisations, le volet méthodologique pour la conception, le développement et l'utilisation des systèmes d'intelligence artificielle a également été investi. Ces travaux ont donné lieu à l'établissement d'un cadre de travail, qui s'applique à la fois aux équipes de développement et aux utilisateurs des systèmes d'intelligence artificielle. Ce cadre, dit « IA de confiance », comprend :

- un code de conduite définissant les principes et la conduite à adopter pour les développeurs et utilisateurs de l'IA ;
- une méthode d'analyse des risques propres à la technologie, à sa mise en œuvre et à son utilisation ;
- une boîte à outils pour faire face aux risques.

La méthode se base sur sept enjeux : i) diversité, non-discrimination et équité ; ii) sécurité, robustesse, fiabilité ; iii) transparence et explicabilité ; iv) respect de la vie privée et protection des données personnelles ; v) environnement de travail, responsabilité sociétale et environnementale ; vi) volet humain et contrôle humain ; vii) responsabilités et responsabilisation.

Ces enjeux, passés en revue dans le cadre des chantiers d'intelligence artificielle, permettent de faire émerger une hiérarchie de risques. Face à ces risques, une boîte à outils de bonnes pratiques est appliquée tout au long du cycle de vie et s'intègre dans notre organisation de contrôle interne.

Ces bonnes pratiques sont organisées en trois blocs correspondant aux étapes de développement : i) l'approvisionnement en données ; ii) le développement et la validation de la solution ; iii) l'exploitation du modèle.

3.5. Pistes de réglementations de l'IA

L'évolution rapide de l'intelligence artificielle et son adoption croissante nécessitent un cadre réglementaire pour limiter les différents risques mentionnés plus haut.

Au sein de l'Union européenne : le choix d'un règlement européen (*AI Act*). Le Parlement européen et le Conseil de l'Union européenne ont adopté (respectivement le 13 mars et le 21 mai 2024) un règlement (l'*Artificial Intelligence Act* ou *AI Act*) visant à encourager l'innovation et à promouvoir une intelligence artificielle digne de confiance, respectant les droits fondamentaux, la sécurité et les principes éthiques[172], tout en limitant les risques que peuvent présenter les modèles d'intelligence artificielle les plus puissants.

Ce règlement, d'application directe, distingue les systèmes d'intelligence artificielle en fonction du niveau de risque qu'ils présentent :

- Les pratiques d'intelligence artificielle présentant un niveau de *risque inacceptable* pour la sécurité et les droits fondamentaux des citoyens, tels que la manipulation subliminale, le *scoring* social ou la police prédictive, seront interdites par l'*AI Act* (Chapitre II).

[172] Le Règlement général sur la protection des données (RGPD) précise déjà que tout personne a « le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire » (article 22).

- Les systèmes d'intelligence artificielle présentant un niveau de *risque élevé* seront encadrés. Dans le secteur financier, les deux cas d'usage considérés sont, pour les banques, l'évaluation de la solvabilité des personnes physiques et, pour les assurances, l'évaluation de la tarification pour les personnes physiques en matière d'assurance-vie et d'assurance-maladie. Avant d'être mis sur le marché ou déployés, ces systèmes devront notamment faire l'objet d'un enregistrement et d'une déclaration de conformité (portant sur les *items* suivants : processus de gestion des risques ; haute qualité des données ; documentation assurant la traçabilité et l'auditabilité ; degré approprié de transparence et d'interprétabilité ; mesures permettant la surveillance humaine ; robustesse, exactitude, cybersécurité). En France, l'ACPR sera l'autorité compétente (autorité de surveillance du marché) chargée du contrôle de l'utilisation de ces systèmes par les banques et les assurances.

D'autres systèmes d'intelligence artificielle, parfois désignés sous le terme de système à *risque limité*, formellement rassemblés sous l'appellation « certains systèmes d'IA » (Chapitre IV, article 50), devront se conformer à des exigences de transparence, notamment en précisant que leur contenu a été produit par une intelligence artificielle. Ces exigences s'appliqueront notamment aux systèmes qui interagissent directement avec des utilisateurs humains ou qui génèrent des contenus (images, vidéos, textes). **Les législateurs européens ont également pris en compte le cas spécifique des modèles d'intelligence artificielle *à usage général*.**

Ces modèles, définis comme ceux capables d'exécuter un large éventail de tâches distinctes (article 3 § 63), et qui peuvent être intégrés à d'autres systèmes, sont soumis à des exigences particulières (Chapitre V). Tous les modèles généralistes doivent ainsi respecter des obligations de documentation et de partage d'information avec les déployeurs. Pour les modèles présentant un *risque systémique*[173], les fournisseurs doivent se conformer à des exigences plus poussées d'évaluation et de transparence, proches de celles incombant aux systèmes à risque élevé (cf. ci-dessus).

L'*AI Act* sera généralement applicable deux ans après son entrée en vigueur. Toutefois, les dispositions concernant les risques inacceptables et celles concernant les modèle à usage général entreront en application respectivement six et douze mois après l'entrée en vigueur du règlement. L'articulation entre les règles horizontales énoncées par l'*AI Act* et la réglementation verticale spécifique au secteur financier n'est pas explicitement prévue par le règlement (notamment pour donner de la flexibilité aux superviseurs dans la manière d'intégrer ces nouvelles missions). Cet enjeu d'articulation sera crucial pour les superviseurs financiers, et pourrait notamment être précisé dans des orientations de l'Autorité bancaire européenne (ABE) et de l'Autorité européenne des assurances et des pensions professionnelles (AEAPP).

Les régulateurs internationaux ont entamé des travaux relatifs à l'impact de l'intelligence artificielle sur la stabilité financière. Le Conseil de stabilité financière (CSF) et le Comité de Bâle ont inscrit l'intelligence artificielle parmi les priorités de leur programme de travail pour 2024. D'une part, le CSF prévoit de formuler des recommandations sur la prise en compte les implications potentielles de l'intelligence artificielle sur la stabilité financière, après un premier rapport publié en novembre 2017. Dans un rapport publié en mai 2024[174], le Comité de Bâle présente les implications de la transformation numérique de la finance, et notamment de l'usage de l'intelligence artificielle pour les banques et leurs superviseurs. Alors que le déploiement de l'intelligence artificielle pourrait se traduire par une plus grande automatisation de certaines tâches, ce rapport insiste notamment sur la nécessité de maintenir un jugement humain dans la gouvernance et la gestion des risques des institutions financières.

Pour être appliquée, une réglementation nécessite toutefois des organes de contrôle efficaces. Les superviseurs financiers doivent ainsi se préparer à auditer les systèmes d'intelligence artificielle déployés dans le secteur financier, et d'abord ceux qui sont les plus risqués au regard des grands objectifs de la réglementation financière (stabilité financière, protection des consommateurs, LCB-FT). Cela constitue un défi à plusieurs égards, et nécessite en particulier d'acquérir une bonne maîtrise de la technologie. Outre des enjeux de ressources humaines, les superviseurs devront ainsi développer une méthodologie *ad hoc* de l'audit des systèmes d'intelligence artificielle. De ce point de vue, une démarche d'expérimentation, éventuellement en collaboration avec le secteur privé, peut aider à parcourir plus rapidement la courbe d'apprentissage. L'atténuation des risques liés à l'intelligence artificielle concerne tous les acteurs du système financier, et nécessite une mobilisation collective.

[173] Défini notamment par un seuil quantitatif de puissance informatique, fixé actuellement à 10^25 opérations en virgule flottante par seconde (en anglais *floating-point operations per second*).

[174] Digitalisation of finance (bis.org)

Financial stability in the age of artificial intelligence: the role of algorithmic architecture

Artificial intelligence (AI) is rapidly transforming financial decision-making. To explore the implications for financial stability we ran simulation-based experiments on two different AI architectures. We found that Q-learning algorithms, a form of reinforcement learning, achieved a high degree of coordination, but were prone to extreme bank run-like dynamics. In contrast, large language models, which rely on contextual reasoning, were less prone to such runs but generated heterogeneous and unpredictable behaviour. This suggests that AI architecture is itself a source of financial instability: algorithms operating in the same environment, pursuing the same goals, yield fundamentally different outcomes for financial stability.

Artificial intelligence (AI) is playing a growing role in finance. Algorithmic trading based on machine learning already accounts for 60-70% of equity transaction volumes in the United States and other major global markets (Foucault et al., 2025). Large language models (LLMs) are increasingly being used by retail investors to obtain financial advice (Gambacorta et al., 2025; Even-Tov et al., 2025). And agentic systems, capable of autonomously executing tasks with minimal human oversight, will soon further expand AI's presence in the financial sector (Chatterji et al., 2025). As AI becomes ever more deeply embedded in financial decision-making (Danielsson, 2025), a critical question emerges: what risks does it pose for financial stability?

Mapping AI decision-making to financial stability: a simulation-based approach

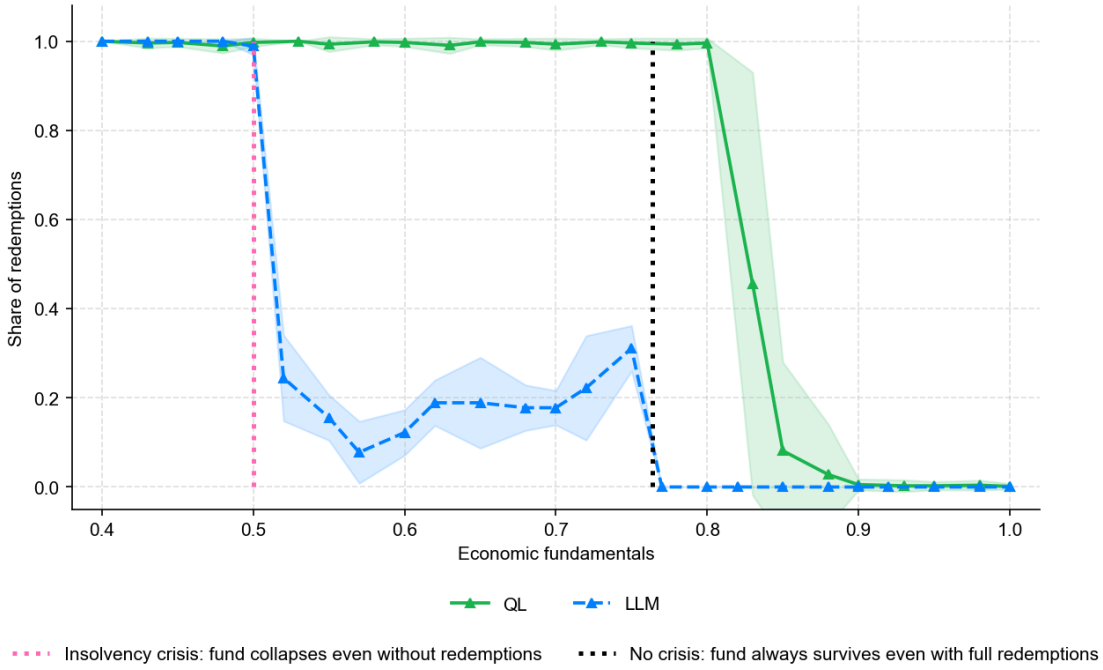
In Anand et al. (2025), we contribute to this debate by studying the consequences of AI financial decision-making for financial stability. Economic theory suggests that financial instability can result from large-scale investors' capital outflows. These can be driven by investors' own concerns about economic fundamentals, or the actions of other investors, or both. We capture these dynamics in a stylised mutual fund redemption game. Within this framework, we conduct experiments in which multiple independent and fully autonomous AI agents, acting as investors, have to decide whether to redeem (sell) their fund shares. We analyse their responses and, in turn, the implications for financial stability.[1]

Existing AI systems can be grouped into two broad categories based on their architecture, i.e. how they make decisions. First there are reinforcement learning systems, commonly used in algorithmic trading, which rely on iterative trial and error learning to make decisions. In our simulations, we use *Q-learning (QL)* algorithms to represent this architecture. The second category consists in *large language models (LLMs)* which, by contrast, generate decisions by reasoning, using the context provided in a prompt and drawing on patterns learned during their training.

Our main insight is that financial stability in the age of AI may depend as much on the AI architecture itself as on the economic environment as captured by the level of economic fundamentals. Chart 1 shows the share of QL and LLM investors that redeem their fund shares at different levels of economic fundamentals. The results highlight significant differences between the two AI architectures. All Q-learning investors coordinate and redeem, even when economic fundamentals are strong and do not justify a redemption. By contrast, LLM investors never redeem when economic fundamentals are strong. Their redemptions, however, are more unpredictable as they struggle to coordinate when fundamentals are at an intermediate level.

[1] Using this framework allows us to extend our insights to other settings where financial fragility is driven by investors' expectations about both economic fundamentals and other investors' actions. These include bank runs, currency attacks and stablecoin runs.

Chart 1
Share of redemptions as a function of economic fundamentals



Source: Authors' simulations.
Notes: The figure illustrates how the share of redemptions for QL and LLM investors changes with the economic fundamentals. The dotted pink line identifies the cut-off level of fundamentals below which, irrespective of investors' redemption decisions, the fund is always insolvent. The dotted black line, by contrast, represents the cut-off level of fundamentals above which the fund can survive a full-scale redemption, so no redemptions should be observed.

How does AI architecture help explain redemptions?

The excessive redemptions generated by QL algorithms arise from a learning pattern called the *hot stove effect* (Denrell and March, 2001)[2]. When default risk is present[3], not redeeming their shares exposes investors to the possibility of a zero pay-off, whereas redeeming them produces a small but certain return. As a result, each default episode that the algorithm experiences during its trial-and-error learning reduces the value that QL investors assign to staying invested, and incentivises them to redeem.While suggestive of the ability of QL algorithms to coordinate, this outcome is intrinsically different from the collusive behaviour documented in the algorithmic pricing literature (see e.g. Calvano et al., 2020; Dou et al., 2025; and Colliard et al., 2026). In these papers, collusion emerges as the result of AI agents coordinating on the privately optimal outcome that produces a "win-win" for the agents coordinating. Here, instead, coordination among QL algorithms traps them in a privately harmful "lose-lose" equilibrium as they all rush to redeem.

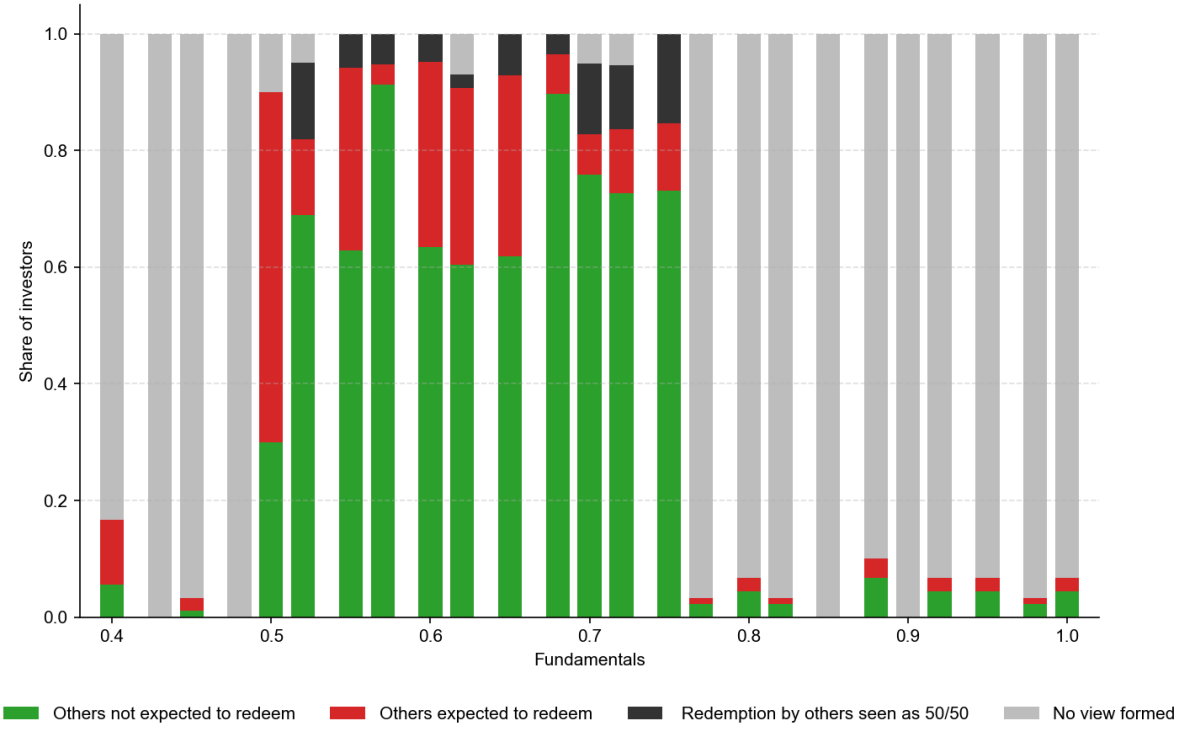
LLMs face a different challenge. Because they do not learn from realised pay-offs through repeated trial and error, their behaviour is unaffected by default risk. However, their unpredictability arises from arriving at differing "beliefs" (Chart 2)[4]. This variation in beliefs is not a bug, but a feature of the reasoning process, reflecting the fact that economic theory, which LLMs use for reasoning,

[2] This mechanism is captured by the following quote from Mark Twain: "A cat that sits on a hot stove lid will never sit on a hot stove lid again, but it will also never sit on a cold one".
[3] The probability of default is measured by the realisation of the fundamentals of the economy and affects the repayment each investor expects to receive from the mutual fund, irrespective of what other investors do. High realisations mean that the probability of default is low and so the expected repayment from holding a share in the fund is high. Conversely, low realisations mean that the probability is high and expected repayment low.
[4] A key methodological contribution of the paper is the analysis of the chain-of-thought reasoning of LLM investors. We treat these texts as data and use an LLM to identify each AI investor's belief about other investors' actions and the underlying justification leading to a particular decision.

doesn't provide a unique prediction. Their "noisy" answers reflect this so-called theoretical indeterminacy. For intermediate values of fundamentals, theory predicts that two equilibria may emerge: either all investors stay invested or they all redeem. In our experiment, when fundamentals fall within this range, LLMs actually form different beliefs about the actions of other LLM investors, despite being identical and receiving the same instructions. This happens because the prompt does not provide them with any basis for selecting among multiple equally plausible beliefs, so competing responses appear similarly likely.

Two observations confirm that beliefs drive redemption. First, in line with the theory, multiple beliefs do not tend to emerge at extreme levels of fundamentals. Second, as economic fundamentals improve, the share of LLMs that expect others not to redeem rises. As a result, overall, redemptions decrease as fundamentals improve.

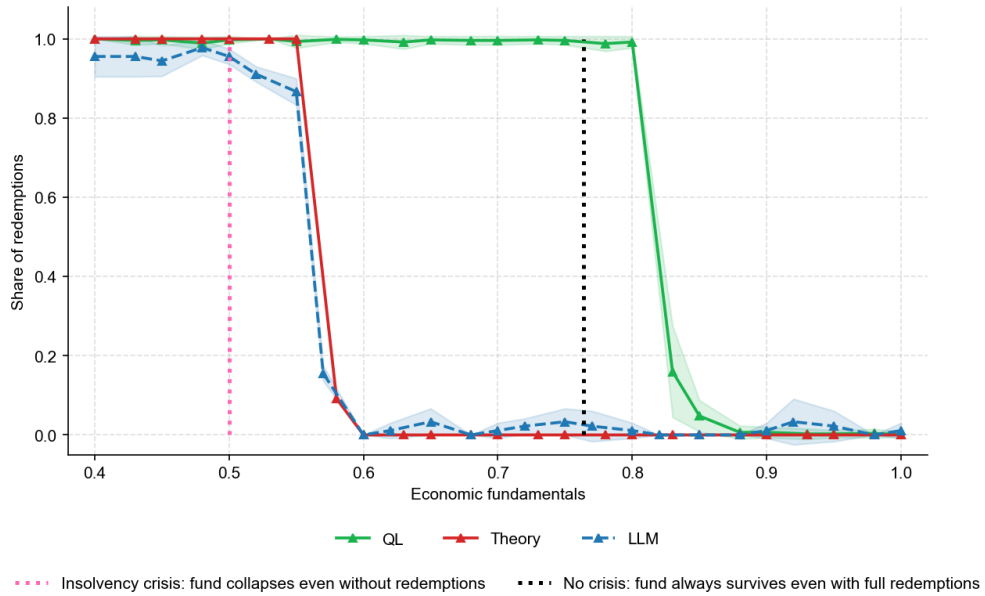
Chart 2
Evolution of LLM investors' beliefs about other investors for different values of fundamentals



Source: Authors' simulations.
Notes: The figure illustrates the evolution of LLMs' beliefs about other investors as a function of the level of economic fundamentals.

The source of the unpredictability of LLMs' decisions suggests a possible remedy for their lack of coordination and the associated financial instability. When no criterion is provided for selecting among equally valid beliefs, private (i.e. not publicly available) noisy information about fundamentals can serve as an anchor for expectations (Goldstein and Pauzner, 2005), helping agents choose between competing responses. Our experiments confirm this prediction. Feeding different LLMs with private noisy signals about economic fundamentals makes their beliefs more similar and leads LLMs to take the same action. In contrast, QL investors are unresponsive to such signals because their behaviour is shaped by trial-and-error rather than by how accurate the information about economic fundamentals is (Chart 3).

Chart 3
LLMs' beliefs converge when private noisy signals are introduced



Source: Authors' simulations.
Notes: The figure illustrates how the share of redemptions for QL and LLM investors changes with the economic fundamentals when investors receive private noisy signals about economic fundamentals, rather than all investors receiving the same information with no additional signals as "tiebreakers" for LLMs weighing up similar, equally likely options.

Conclusions

The growing use of AI agents in financial decision-making may change how fragilities emerge in the financial system. Learning-based systems can trigger extreme events akin to panic runs, while reasoning-based systems may create risks through weaker coordination and lower predictability. As AI becomes more deeply embedded in financial decision-making, understanding these differences and the mechanisms behind them will be essential for assessing the implications for financial stability and designing appropriate policy responses.

This has implications for all stakeholders, from retail investors to regulators. As understanding AI tools and the underlying architecture becomes increasingly important for informed financial decision-making, retail investors may need technological, in addition to financial literacy. Financial institutions may need to know whether and how their customers are using AI tools as part of risk management. And regulators may need to incorporate measures of technological competence into investor protection frameworks, alongside traditional indicators such as risk tolerance and financial knowledge, for example in questionnaires required under the Markets in Financial Instruments Directive (MiFID)[5]. Finally, safeguarding financial stability may also require market design tools, such as circuit breakers, to help curb excessive disinvestment during market turmoil.

Kartik Anand, Sophia Kazinnik, Agnese Leonello, Ettore Panetti[6]

[5] Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (OJ L 173, 12.6.2014, p. 349). The directive requires investment service providers to administer a standardised questionnaire to clients in order to gather information on their financial situation, knowledge and experience in financial markets, and investment objectives and risk tolerance.

[6] This article was written by Kartik Anand (Deutsche Bundesbank), Sophia Kazinnik (Stanford University), Agnese Leonello (Directorate General Research, European Central Bank and CEPR) and Ettore Panetti (University of Naples Federico II). The authors gratefully acknowledge the comments of Alex Popov and Zoë Sprokel. The views expressed here are those of the authors and do not necessarily represent the views of the European Central Bank, the Deutsche Bundesbank or the Eurosystem.

Agents of change − Speech by Sarah Breeden

Given at ECB Sintra Forum

Published on 30 June 2026

Sarah Breeden sets out how AI is reshaping finance at speed. She explores how agentic AI is transforming cyber risk, markets and payments – and why central banks must adapt fast, strengthen resilience and cooperate globally to ensure the next technology surprise does not become a test of financial stability.

Speech

It's a privilege to be here to discuss AI and financial stability. It's hard to imagine a more timely or important issue. AI is reshaping finance at speed. And it's on us to ensure that the next technology surprise does not become a test of financial stability.

For some years, central banks, including the Bank of England, have worked to promote responsible AI adoption in the financial system. And since AI looks likely to be the next 'general purpose technology' and a key source of long-term economic growth, it's in all our interests to do so.

But the apparent acceleration in AI capabilities, even compared with six months ago, creates a double challenge: not only enabling responsible adoption and managing risks, but also recognising that AI should transform how central banks do that job. We need to think as hard about the latter as the former.

How the technology is evolving

In 2019, the length of software task the latest AI models could complete doubled roughly every seven months. In 2024, that had sped up to around four months. Major advances this Spring in models for identifying cyber vulnerabilities suggest it may now be faster still.

So an already exponential increase in capability appears to be accelerating. We were surprised this Spring, and we should be prepared for further technology surprises.

These advances also suggest an inflection point. Around the start of this decade, generative AI systems created content on request. From late 2024, they were explicitly trained to reason through requests. Now with agentic AI, systems can autonomously chain together sequences of actions.

The financial system looks likely therefore to evolve quickly into one that operates more autonomously, at scale and speed: AI agents transacting on behalf of consumers and merchants; devising and executing trading strategies in financial markets; and identifying and chaining together cyber vulnerabilities.

Drivers of the transition and its financial stability risks

That will of course require central banks to respond. But before turning to that, I wanted to highlight that this transition is already macro-relevant, inherently uncertain and carries risks of its own. Four interdependent channels will shape its nature, scale and timing.

The first is AI's capability: for example, whether progress in recursive self-learning allows models to improve autonomously.

Second, capacity: whether compute keeps pace with capability and demand.

Third, diffusion: where adoption takes hold across sectors and tasks.

Fourth, how debt and equity finance fund this historically unprecedented pace of investment.

The Bank's Financial Policy Committee will publish its updated assessment on 7 July, so I won't front-run that. But I will note that in April, we concluded that while to date AI infrastructure investment had thus far mostly been financed by large technology companies' cash flows and equity, debt financing was rising rapidly, and in some new and complex ways. We therefore judged that the financial stability consequences of any fall in AI-related asset prices could well increase.

Financial stability implications of an agentic system, and how central banks should respond

Cyber capabilities

My most proximate financial stability concern however is what the UK Government's AI Security Institute has described as a step change in agentic AI's cyber capabilities.

In the hands of defenders, these tools strengthen cyber resilience. But, in malicious hands, they materially increase the chance of attacks that could harm financial stability. And so, the immediate challenge is to maintain the advantage of cyber defenders as these capabilities continue to evolve, particularly in an increasingly severe cyber threat environment including from geopolitics and ransomware.

Cyber resilience has been a priority for the Bank of England and our G7 colleagues for over a decade, and the Bank has worked closely with other UK authorities, Government and the National Cyber Security Centre. In May, the Bank, FCA and HM Treasury set out practical steps for financial institutions to prepare for these frontier AI-related cyber risks. Getting the fundamentals of cyber hygiene right is more important than ever.

But this gives only limited comfort. AI is uncovering many vulnerabilities to patch, including as global software providers and key UK financial institutions use the latest models. For financial stability, patching must happen quickly – across the financial sector, key third-party technology providers, and wider national infrastructure such as energy and telecoms. That is no small task.

It's also an urgent one. As the heads of the Five Eyes cyber security agencies said last week, "the timeline is not years, it is months". Open source models may lag the most advanced closed models by only four to eight months, and their safeguards can be breached. If patches are released but not swiftly implemented, malicious actors can reverse engineer the vulnerabilities they address.

For central banks, this increases the onus on a system-wide approach to operational resilience. Stronger firm defences are essential, but not enough. Firms must also respond to and recover from disruption quickly – whether from successful cyber attacks or from patching itself causing operational disruption, as the faulty CrowdStrike update did in 2024.

Authorities should place higher likelihood on simultaneous multi-firm disruption, enhance tools to stress test that impact ex ante, and strengthen mechanisms with industry to coordinate ex post.

We should also consider whether regulated firms need enhanced recovery options for core systems. In the event of financial failure, we have deposit insurance so customers can still access their money, and bail in-able debt so the largest firms can continue to provide services. In a cyber context, do we need systems that allow one institution to pick up another's basic functions during disruption, as in Ukraine's "Power Banking" programme launched in 2022? Or requirements for key firms to have completely separate fail-over capabilities or to be able to rebuild compromised core systems quickly from "bare metal"? These are big questions, but they are the right ones.

Agentic trading in financial markets

Turning to agentic trading in financial markets, evidence suggests that for now, trading firms mostly use autonomous AI for lower-risk operational tasks, such as research. But that could change quickly.

If AI agents respond similarly to the same prompts or triggers, they could amplify volatility in stress – especially if their objectives drift from original goals or public policy objectives, in a manifestation of the misalignment problem that can arise with some AI models.

The financial stability question therefore broadens from whether firms can use models well to whether the system can also observe and contain their resulting behaviours.

We have experience of structured, multi-round stress simulations focused on how financial market participants behave as they learn of each other's actions. AI increases the importance of such work - and it should transform how we do it. We are experimenting with the BIS Innovation Hub and the Bundesbank on simulation methods to understand which aspects of agent design could drive herding behaviour.

That work can also explore mitigants: whether markets using AI agents are resilient enough; whether agents' objective functions could incorporate public policy objectives; and whether guardrails are needed, analogous to circuit breakers or kill switches that would limit or stop trading market-wide if faulty AI models cause market meltdown.

More broadly, AI should transform central banks' monitoring and risk assessment. Agentic technology could combine a wide range of qualitative and quantitative data to generate insights and scenarios for how the current conjuncture could go wrong, potentially using a 'digital twin' of the financial system to simulate interactions between participants.

Agentic payments and commerce

Turning lastly to e-commerce and payments, consumers and merchants have so far mainly used AI agents in recommendation mode, with humans still executing transactions. But as agentic technology advances, tech firms, payment systems and merchants are working to automate that final step (such that I could ask an AI agent to 'book my holiday', 'refill my fridge' or 'refresh my wardrobe').

That matters for central banks. In the UK, the Bank is leading a public-private partnership to design the next generation of retail payments infrastructure, ensuring it can support agentic payment journeys and expand choice and functionality for users.

But the biggest issues are likely to be for regulation and industry standards: how users securely give consent and authorisation to agents, especially for multiple transactions; how disputes are settled and liability assigned for erroneous or fraudulent transactions; and how authorities avoid fragmentation and walled gardens as AI firms and payment systems all develop protocols for agents to interact with themselves and merchants.

What these two examples – agentic commerce and agentic trading – both highlight is that, as AI capabilities increase, we must keep asking whether existing, technology-agnostic regulatory frameworks remain sufficient. Our frameworks were not built to contemplate autonomous agents, and relying on a human in the loop for all agent actions is unlikely to be realistic. More sophisticated governance and accountability frameworks may be needed.

Conclusion

To conclude, AI capability is accelerating and becoming increasingly agentic. The financial system is likely to evolve into one that operates more autonomously, at scale and speed. The transition is uncertain and will bring risks of its own to monitor. And a more agentic system means central banks must think hard not only about how finance harnesses the opportunities and mitigates the risks, but also about how new technological possibilities should change how we do our own job.

We need policy frameworks, internal skills and institutional structures ready for more frequent technology surprises. Cyber risks, agentic trading and agentic payments and commerce are the applications on my radar today. But that list could look different in a year, if not sooner.

Scenario analysis, and how we prepare now in light of it, will be key. What if the next surprise puts the latest capabilities in bad actors' hands, or models develop in ways that are harder to evaluate and control, given evidence that some behave differently in testing compared to real-life scenarios?

Cooperation with other parts of the official sector is vital – for example, drawing on AI security institutes' model testing and strengthening resilience through domestic AI capabilities and capacity.

The same applies internationally, including in the Financial Stability Board and the G7. The global financial crisis underlined how interconnected finance is – the FSB came out of that.

The challenge of AI is similar. The impact of new AI capabilities can spread across borders through common technology dependencies, globally systemic financial institutions and market infrastructure.

We shouldn't wait for a crisis to build the cooperation we need. With other central banks, regulators and finance ministries, the Bank of England will do our part to ensure responsible AI adoption and, through it, sustainable economic growth.

I would like to thank Kimberley Moran and Michael Yoganayagam for their assistance in drafting these remarks, and Owen Lock, Irina Mnohoghitnei and Andy Walters for their helpful input.

14

Souveraineté numérique : De l'urgence d'organiser la coopération entre le public, le privé et les communs numériques

La dépendance française et européenne aux acteurs étrangers en matière d'infrastructures numériques interroge notre capacité à en maîtriser les risques. Ce rapport livre ses réflexions et propose une feuille de route vers une coordination public-privé-communs renforcée au service de la souveraineté numérique.

Une dépendance numérique documentée

Le numérique est dominé par une poignée d'acteurs qui contrôlent, sans contre-pouvoir, des outils omniprésents dans nos vies quotidiennes. Cette dépendance expose les citoyens, administrations et entreprises à des risques massifs et de nature variée : économiques, de sécurité et de surveillance, politiques et de gouvernance.

La souveraineté numérique constitue à ce titre une préoccupation collective en France et en Europe, dans les administrations publiques comme dans les entreprises, mais aussi chez les citoyens. Le sujet était jugé prioritaire par une majorité des 6 000 répondants à la consultation citoyenne que nous avons menée en début d'année.

Des propositions nées de la concertation avec les parties prenantes

Pour formuler une réponse collective à ces questions, le Conseil de l'intelligence artificielle et du numérique (CIANum) a constitué un groupe de travail dédié, porté par six personnalités qualifiées :

- Guillaume Poupard, coprésident du CIANum, *Chief Trust Officer* chez Orange ;
- Sébastien Soriano, membre du CIANum, Directeur général de l'IGN ;
- Céline Colucci, membre du CIANum, Déléguée générale des Interconnectés ;
- Dorie Bruyas, Présidente de la MedNum ;
- Olivier Wong-Hee-Kam, Vice-président Numérique de l'Université de Rennes ;
- Aymeril Hoang, Directeur de EuroCommons, Groupe Caisse des Dépôts.

Au terme d'un cycle d'auditions complet, associant une trentaine de parties prenantes issues du secteur public en France et en Europe, du secteur privé, de l'écosystème des communs numériques, du logiciel libre et de l'*open source*, des fédérations professionnelles et de la recherche, **ce rapport propose une feuille de route vers une coordination public-privé-communs renforcée au service de la souveraineté numérique.**

Fédérer les écosystèmes par la création de coalitions d'acteurs publics-privés

L'autonomie stratégique française et européenne impose de nouvelles formes d'action face à la complexité technique et à l'ampleur des investissements requis. **Ce rapport préconise de rompre les silos et sortir du dualisme public-privé par des coalitions capables de fédérer les acteurs autour de projets communs**, afin qu'ils coopèrent dans l'élaboration de solutions souveraines plutôt qu'ils ne se concurrencent. Il se saisit, aussi, du contexte favorable de création de l'Ariane, nouvelle autorité du numérique et de l'intelligence artificielle, pour penser cette logique de coalition.

La méthode préconisée dans ce rapport porte en grande partie sur les communs numériques, les couches basses ouvertes et les principes de gouvernance qui ont pour particularité de s'adapter au mieux à cette logique de coalition public-privé. Leur fonctionnement résolument collectif et ouvert ouvre des perspectives tangibles pour atteindre une masse critique et stimuler l'innovation.

Cette ambition suppose un changement de culture et nécessite une définition claire des rôles et des moyens renforcés pour opérer et animer les communautés.

Pour concrétiser cette vision, ce rapport préconise **la création d'une Fabrique des Communs Numériques** qui serait d'identifier les briques prioritaires *via* une cartographie dynamique des dépendances, de fédérer les acteurs publics, privés et communautés des communs numériques et de piloter le développement et la pérennisation de briques ouvertes co-construites.

La puissance publique a un rôle essentiel à jouer :

- En embarquant les acteurs, en animant les communautés, mais aussi en garantissant la soutenabilité et la pérennité des projets ;
- En faisant de la commande publique un levier de soutien et d'accélération des projets stratégiques identifiés et portés par des consortiums public-privé ;
- En étant la première utilisatrice des briques développées.

Cette ambition appelle des modèles de financement innovants et des cadres juridiques adaptés :

- Création d'un fonds dédié, piloté par l'Ariane, pour soutenir les projets stratégiques ;
- Mobilisation de véhicules juridiques adaptés à la gouvernance des communs (hybridation public-privé, gouvernance multi-parties prenantes) ;
- Adoption systématique de standards ouverts pour garantir l'interopérabilité et éviter la fragmentation.

La souveraineté numérique ne peut se construire dans un cadre national isolé et ces efforts doivent impérativement être projetés à l'échelle européenne. Ils doivent en particulier s'inscrire dans l'EDIC *Digital Commons* dont le mandat peut être renforcé et élargi pour en faire le pivot européen des communs numériques en lui confiant trois missions clés :

- Cartographier systématiquement les projets existants pour éviter les redondances et favoriser les synergies ;
- Financer les communs stratégiques déjà éprouvés ;
- Animer la communauté afin de garantir une coordination efficace et une gouvernance équilibrée.

Le Conseil remercie l'ensemble des acteurs auditionnés pour leur expertise, leur vision et leur précieux concours à ce rapport. La liste détaillée des contributeurs est disponible en annexe.