

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 financière¹²⁵, 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 européenne¹²⁶ 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 1940-1960¹²⁷ avec la création d'un premier modèle mathématique de réseau de neurones en 1943¹²⁸ 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 Turing¹²⁹. 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 \(coe.int\)](#)

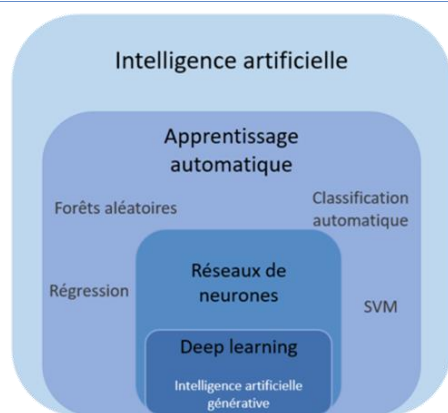
¹²⁸ MCCULLOCH, Warren S. PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 1943, vol. 5, p. 115-133.

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*.

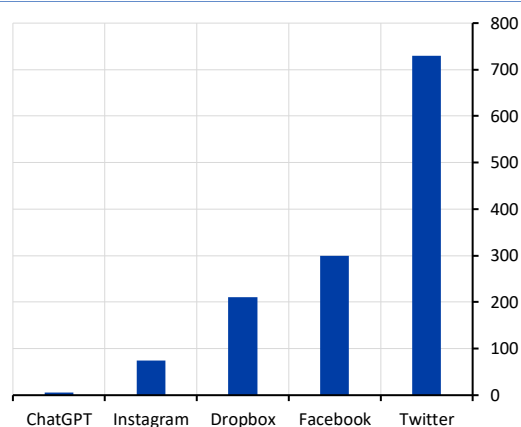
¹³⁰ L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle, qui se concentre sur le développement de programmes informatiques conçus pour apprendre au fur et à mesure de leur développement et de leur utilisation. L'apprentissage profond est une branche de l'apprentissage automatique 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 sondage¹³¹, 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 récente¹³², 35 % des entreprises déclarent utiliser des technologies liées à l'intelligence artificielle¹³³.

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 événements 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 années¹³⁴. 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.

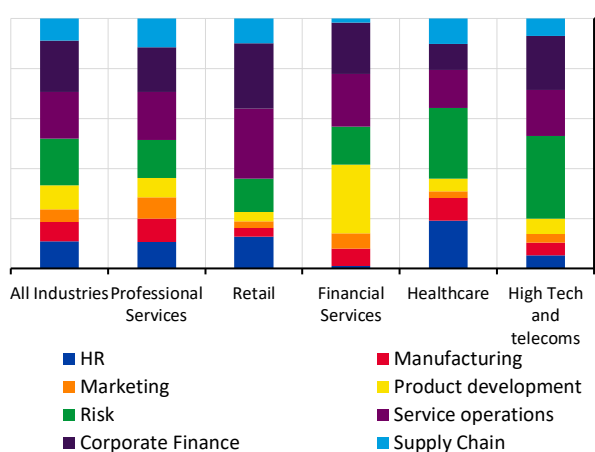
¹³³ Diagnostic structurel de la France 2023 (note d23-124)

¹³⁴ Generative AI: Differentiating disruptors from the disrupted | MIT Technology Review

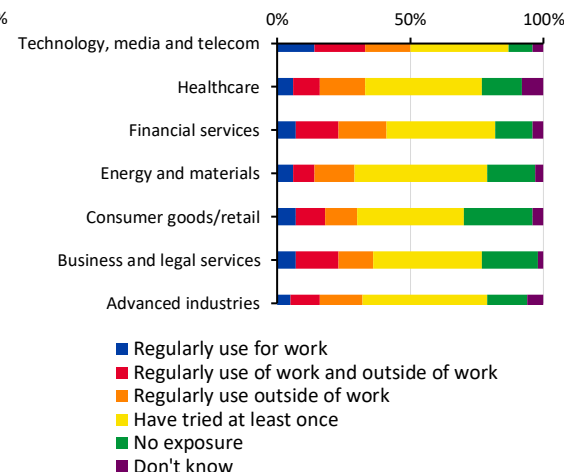
entreprises citent les enjeux de confidentialité des données, de confiance, de transparence et le manque de compétences¹³⁵.

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 qualifiés¹³⁶. 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 place¹³⁷. 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'euros¹³⁸. Un article publié en décembre 2023 par le FMI¹³⁹ 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 identifiées¹⁴⁰ :

- 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 meilleure 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 pratiques¹⁴¹.

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 réglementation 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 conformité¹⁴³.

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 \(moody's.com\)](#)

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 2023¹⁴⁴, une large majorité d'institutions financières anticipe une hausse des applications liées 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 clients¹⁴⁵. 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 langage naturel¹⁴⁶. 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 risques¹⁴⁷. 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 marché¹⁴⁸ 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 2023¹⁴⁹, 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 \(banque-france.fr\)](#)

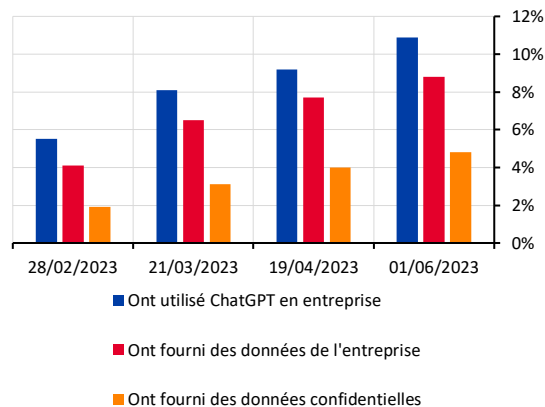
¹⁴⁸ [Written reply to Parliamentary Question on use of artificial intelligence in supervision of financial institutions \(mas.gov.sg\)](#)

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 moindres dans le cas d'outils réservés à l'usage interne dans les entreprises, avec des règles strictes de gestion des données.

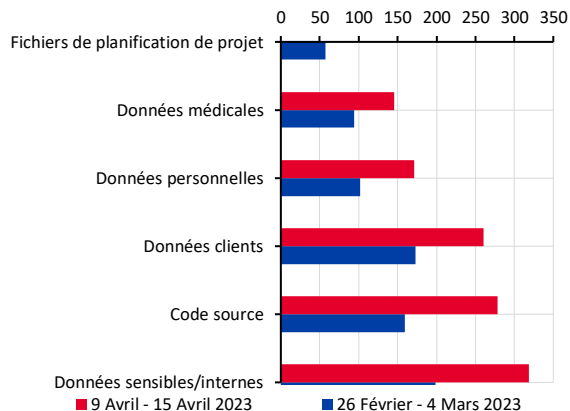
Graphique 3.5 : Adoption de ChatGPT par les salariés

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 erronés¹⁵⁰. 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 obsolètes¹⁵¹.

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 ailleurs¹⁵². 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, Iliia, 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 nouveaux¹⁵³.

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 Google¹⁵⁴. 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 \(bis.org\)](https://www.bis.org/press/pr240101.htm)

¹⁵⁴ [The leading generative AI companies \(iot-analytics.com\)](https://www.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 marchés¹⁵⁵. 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'événements 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 technologie¹⁵⁶ 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 chocs¹⁵⁷. 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 spécialisés¹⁵⁸. 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 formes¹⁵⁹. 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 effectués¹⁶⁰, mais aussi pour étudier les variables prises en compte et le poids assigné à chacune d'entre elles pour l'obtention du résultat¹⁶¹. 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 prudentielles¹⁶².

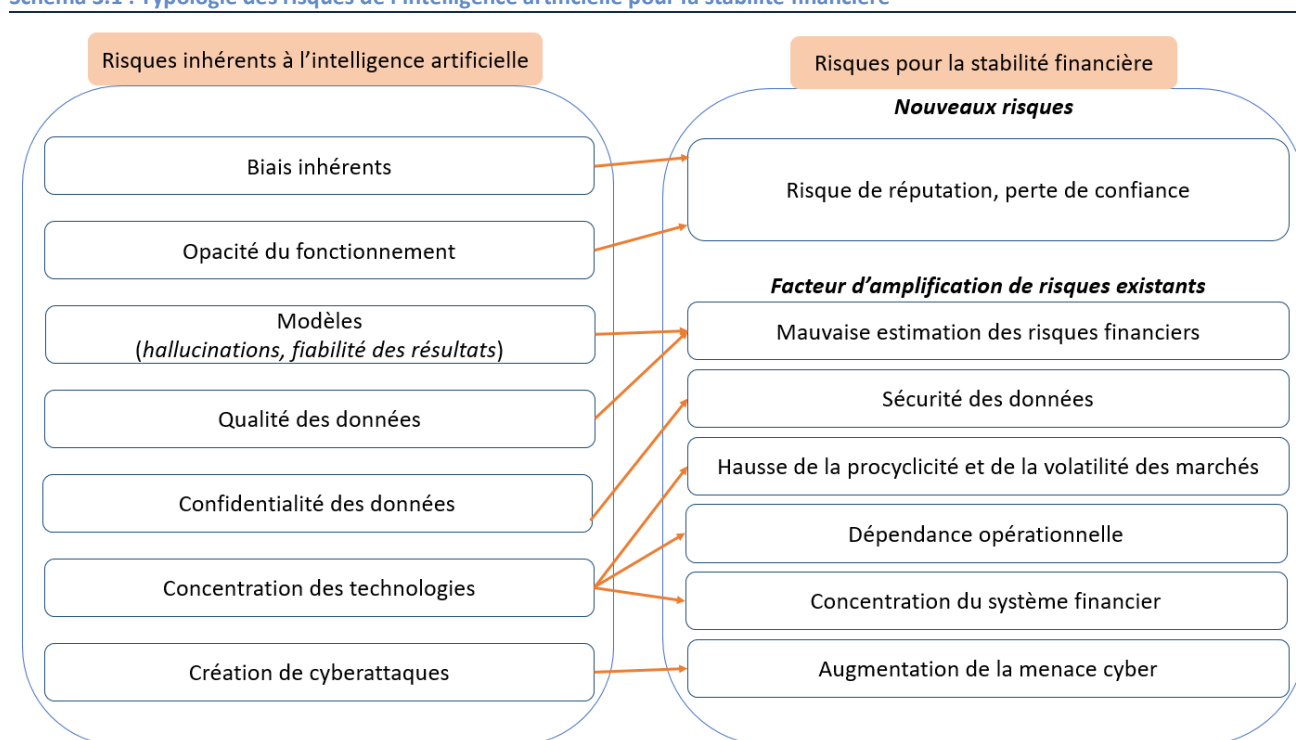
¹⁵⁹ [La motivation du conseil par les robo-advisors : vers un éclairage apporté aux clients ? | ACPR \(banque-france.fr\)](#)

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 sociaux¹⁶³. 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

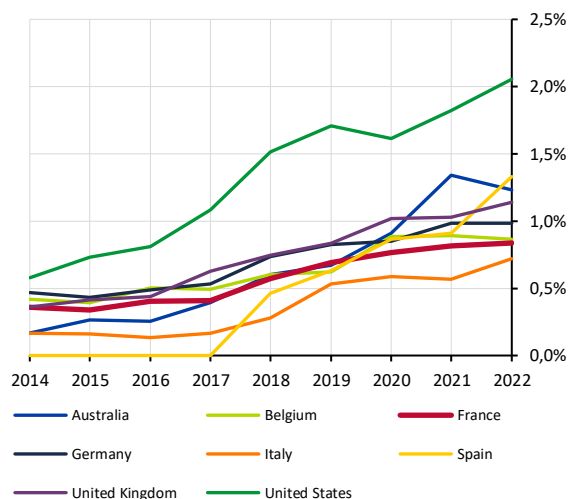
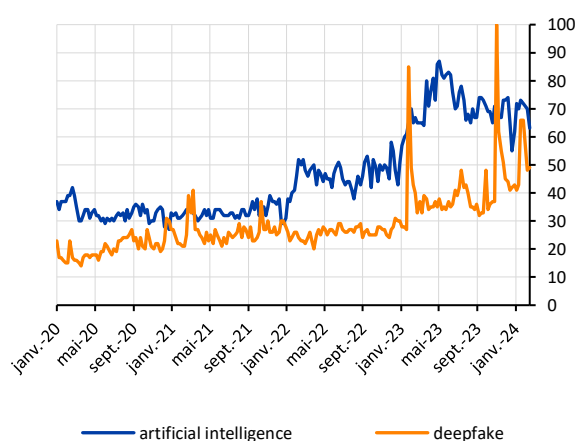
¹⁶³ Social Media as a Bank Run Catalyst by J. Anthony Cookson, Corbin Fox, Javier Gil-Bazo, Juan Felipe Imbet, Christoph Schiller :: SSRN

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

x : axe temporel / y : indice

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 certitude¹⁶⁵. Tous les pays et toutes les fonctions économiques ne sont pas également exposés 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 artificielle¹⁶⁶. 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'OpenAI¹⁶⁸, 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 recherche¹⁶⁹. 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 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 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 industrielle¹⁷⁰ y compris avec des partenaires extérieurs¹⁷¹, permettant de couvrir de nombreux pays avec des méthodes homogènes.

¹⁶⁸ [AI and Compute \(georgetown.edu\)](https://www.georgetown.edu)

¹⁶⁹ [Generative AI's environmental costs are soaring — and mostly secret \(nature.com\)](https://www.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 éthiques¹⁷², 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 toute 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 modèles 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 des implications potentielles de l'intelligence artificielle sur la stabilité financière, après un premier rapport publié en novembre 2017. Dans un rapport publié en mai 2024¹⁷⁴, 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

¹⁷³ Défini notamment par un seuil quantitatif de puissance informatique, fixé actuellement à 10²⁵ opérations en virgule flottante par seconde (en anglais *floating-point operations per second*).

¹⁷⁴ [Digitalisation of finance \(bis.org\)](https://bis.org)

liés à l'intelligence artificielle concerne tous les acteurs du système financier, et nécessite une mobilisation collective.