

# SFI Public Discussion Note

## Le rôle de l'IA dans la transformation des pratiques financières



# Introduction



**Prof. Markus Leippold**  
*SFI Senior Chair et professeur  
d'ingénierie financière à  
l'Université de Zurich*

A travers sa série de *Public Discussion Notes*, le Swiss Finance Institute (SFI) encourage activement des échanges étayés autour de thèmes pertinents pour le secteur financier, les milieux politiques et le monde universitaire. En outre, le SFI diffuse ses résultats par le biais de travaux de recherche, de publications, de Master Classes et de conférences.

Cette *Public Discussion Note* est basée sur les recherches que mon équipe et moi-même avons menées dans les domaines de la durabilité et de l'IA au cours des dernières années. Je tiens à remercier mes coauteurs, Julia Bingler, Chiara Colesanti-Senni, Glen Gostlow, Jingwei Ni, Tobias Schimanski, Dominik Stambach, Saeid Vaghefi et Tingyu Yu, ainsi que tous les assistants de recherche. Ces jeunes chercheurs talentueux et motivés ont joué un rôle déterminant dans l'avancement de ce travail. J'espère que leurs contributions inciteront les futurs chercheurs à continuer à progresser dans ce domaine important.

Cette version est une traduction de la version originale en anglais. La version originale est disponible à l'adresse suivante : <https://sfi.ch>.

"Je pense que l'intelligence à l'ancienne fonctionne plutôt bien" (Munger, 2023).

Depuis l'introduction de *ChatGPT* par *OpenAI* en novembre 2022, l'intelligence artificielle (IA) suscite l'engouement de Wall Street. Pourtant, si l'IA est devenue un sujet incontournable lors de la présentation des résultats trimestriels des entreprises, certains investisseurs éminents demeurent sceptiques et estiment que le battage médiatique autour de l'IA est exagéré.

Le PDG de Berkshire Hathaway, Warren Buffett, l'un des investisseurs les plus avisés au monde, oppose intelligence artificielle et intelligence "à l'ancienne", comme le faisait son comparse de longue date, le regretté Charles Munger. À l'heure où d'autres acteurs se montrent de plus en plus fascinés par les capacités de l'IA, ainsi que par ses progrès rapides et son intégration dans les pratiques des entreprises, Buffett met en avant la valeur pérenne et précieuse de l'intelligence humaine en se posant la question de savoir si l'IA est réellement en mesure de surpasser la profondeur et la nuance de la pensée et de la créativité humaines.

À l'instar de l'époque de la révolution industrielle, la période actuelle marquée par l'essor de l'IA se caractérise par des progrès technologiques rapides et par un fort optimisme concernant ses conséquences sur nos structures sociales, notre économie et notre vie quotidienne. Il y a quelque 200 ans, l'introduction de méthodes de production mécanisées avait marqué un tournant important dans l'histoire de l'humanité. L'avènement de l'IA en augure un autre. Tout en explorant la capacité des machines à créer, raisonner et interagir d'une manière que nous croyions jusqu'à présent réservée aux humains, nous devons être conscients des problèmes en matière d'éthique, de protection de la vie privée et de sécurité qui découleront d'une adoption généralisée de l'IA.

Dans la présente *Public Discussion Note*, j'espère parvenir à trouver un juste équilibre entre exubérance et prudence. En examinant les prévisions optimistes des grandes entreprises technologiques, qui considèrent l'IA comme un élément déterminant pour rester compétitif sur le marché, mais également les perspectives plus sceptiques d'investisseurs tels que Buffett et Munger, je tenterai de démêler les complexités du phénomène actuel de l'IA. Après avoir discuté de la technologie elle-même et de la capacité de l'IA à remodeler le paysage économique, je poursuivrai en abordant les considérations éthiques, les implications sociétales et les défis qui doivent être relevés avant que l'IA puisse être largement généralisée dans le domaine de la finance.

# Une brève histoire de l'IA

## Visions préliminaires (avant 1950) :

Depuis l'Antiquité, les civilisations ont imaginé des formes de vie artificielle, comme en témoignent les mythes de Pygmalion et du Golem. Dans l'histoire de Pygmalion, un sculpteur tombe amoureux d'une statue qu'il a taillée dans la pierre et qui prend vie. De manière similaire, le Golem est une créature mystique faite d'argile qui prend vie afin de servir son créateur. Dans l'Antiquité comme au Moyen-Âge, des inventeurs de diverses cultures ont créé des dispositifs mécaniques imitant les actions humaines ou animales. Ces "automates" reflètent également des visions préliminaires de l'intelligence artificielle.

Des philosophes tels que René Descartes (né en 1596) et Gottfried Leibniz (né en 1646) ont joué un rôle essentiel dans la réflexion sur l'esprit, l'intelligence et la possibilité pour les machines d'imiter le raisonnement humain. Descartes s'est interrogé sur la nature de la pensée et de la conscience, tandis que Leibniz a imaginé un langage universel de raisonnement, connu sous le nom de *calculus ratiocinator*. Ce premier cadre conceptuel, qui représente des énoncés logiques et des raisonnements de manière précise et mathématique, a influencé la logique symbolique et l'informatique moderne.

## Naissance de la discipline (années 1950) :

Les travaux d'Alan Turing dans les années 1940 et 1950, et notamment l'élaboration du test de Turing, ont jeté les bases conceptuelles de l'évaluation de l'intelligence des machines. Le test de Turing est une méthode qui permet de déterminer si une machine peut avoir un comportement intelligent indiscernable de celui d'un être humain. Les travaux antérieurs de Turing ont également établi des principes fondamentaux qui sous-tendent l'informatique moderne. La machine de Turing, par exemple, est un dispositif théorique qui manipule des symboles sur un ruban selon un ensemble de règles, fournissant un modèle simple mais puissant pour comprendre ce que signifie le fait qu'une fonction soit calculable.

La conférence de Dartmouth (1956) a rassemblé des chercheurs qui partageaient la conviction que l'intelligence artificielle était possible. Nous considérons souvent que cette conférence, qui a ouvert la voie à de nouvelles recherches, a marqué la naissance officielle de l'IA en tant que discipline scientifique.

## Une progression en dents de scie (1960-1990) :

Les progrès réels de l'IA n'étant pas à la hauteur des promesses qui l'entouraient, la discipline a subi des périodes marquées par des gels de financement et des baisses d'intérêt. Ces "hivers de l'IA" ont eu lieu principalement entre la fin des années 1970 et le début des années 1980, puis entre la fin des années 1980 et le début des années 1990. Durant ces périodes, un fort scepticisme concernant la capacité de l'IA à tenir ses promesses a entraîné une réduction des investissements et une réévaluation des objectifs.

Malgré ces revers, le développement des systèmes experts a représenté un succès important pour l'IA dans les applications commerciales et industrielles. Ces systèmes étaient capables d'imiter les capacités de prise de décision d'experts humains dans des domaines spécifiques, tels que le diagnostic médical et la prospection minière. Dans le domaine de la finance, les systèmes experts ont été utilisés pour estimer les pertes pour défaut de paiement et ont permis de fournir aux banques et aux institutions financières des outils sophistiqués pour évaluer le risque de crédit et prédire la probabilité que les emprunteurs ne puissent pas rembourser leurs prêts.

## Résurgence et transformation (depuis 1990) :

La résurgence de l'IA a été alimentée de manière significative par les progrès de l'apprentissage automatique (*machine learning*) et par la disponibilité de vastes ensembles de données connus sous le nom de "*big data*". Ces développements ont permis aux systèmes d'IA d'apprendre à partir d'exemples, plutôt que de s'appuyer uniquement sur des règles codées en dur.

L'apprentissage profond (*deep learning*), un type d'apprentissage automatique utilisant des réseaux neuronaux, a permis des avancées significatives dans les domaines de la vision par ordinateur, de la reconnaissance vocale et du traitement du langage naturel, rendant possibles des applications telles que la traduction linguistique en temps réel, les véhicules autonomes et une recherche efficace sur le Web.

# Motivation

Au début des années 1990, lorsque j'étais jeune, la lecture du roman de science-fiction "Golem XIV" de Stanislaw Lem, avait éveillé ma fascination pour les possibilités de l'IA. L'histoire est celle d'un personnage prénommé Golem, une machine d'une intelligence sans précédent qui s'engage dans de profondes discussions avec les plus grands scientifiques du monde. Le concept me semblait purement théorique à l'époque, mais ce roman écrit dans les années 1980 reflète de manière étrangement prémonitoire les interactions actuelles avec l'IA dans des secteurs tels que la finance. L'exploration par Lem des capacités de l'IA, de ses implications éthiques et de son potentiel à coexister avec les humains trouve aujourd'hui un écho tout particulier, à l'heure où des technologies telles que les outils ChatGPT d'OpenAI et Gemini de Google s'intègrent dans nos activités quotidiennes.

Je participe désormais activement aux développements technologiques imaginés par Lem. Les réflexions autour de ce roman soulignent l'importance de considérer avec prudence et clairvoyance le rôle croissant de l'IA dans le secteur financier. C'est le même équilibre délicat que nous devons atteindre entre l'exploitation des avancées technologiques et la prise en compte de leurs implications plus larges. Cette réflexion devient particulièrement importante lorsque nous observons l'impact transformateur de l'IA dans des domaines comme la médecine, la biologie et la météorologie.

Par exemple, le système *AlphaFold*, développé par *Google DeepMind*, a révolutionné la biologie en prédisant avec précision les structures 3D des protéines en se basant uniquement sur leurs séquences d'acides aminés (Jumper et al., 2021). Cette avancée permet de résoudre le problème de longue date du repliement des protéines, d'accélérer la découverte de médicaments et d'améliorer notre

compréhension des processus biologiques en prédisant les structures de l'ensemble des quelque 200 millions de protéines actuellement connues en l'espace de seulement un an. Quand nous savons qu'il faut environ entre quatre et cinq ans à un doctorant pour découvrir la structure d'une seule protéine, cela signifie que cette tâche aurait sinon nécessité des centaines de millions d'années de travail humain! Demis Hassabis et John M. Jumper, tous deux de *Google DeepMind*, ont reçu le prix Nobel de chimie 2024 pour leurs contributions révolutionnaires à la prédiction du repliement des protéines. Dans le même état d'esprit, *GraphCast*, un autre modèle innovant de *Google DeepMind*, a redéfini les prévisions météorologiques en surpassant avec une précision remarquable les méthodes classiques (Cookson, 2023; Lam et al., 2023). En analysant de nombreuses données historiques à l'aide d'un réseau neuronal graphique, *GraphCast* fournit des prévisions de qualité supérieure des conditions météorologiques, de la température aux configurations des vents en passant par les points d'atterrissage des ouragans. Cet outil réduit également de manière significative la consommation d'énergie en ramenant le temps de calcul nécessaire de plusieurs heures à une minute.

Pourrions-nous assister à un "moment" comparable aux cas *AlphaFold* et *GraphCast* dans le secteur financier? En découvrant des modèles et en tirant des enseignements de vastes ensembles de données, l'IA générative a le potentiel de transformer fondamentalement l'analyse financière, l'évaluation des risques et les prévisions économiques. Des outils d'IA puissants pourraient redéfinir notre processus de prise de décision stratégique. Le secteur financier doit donc être prêt et s'adapter afin d'obtenir des résultats significatifs et responsables.



# L'IA générative dans le secteur financier

Avant d'examiner l'impact de l'IA générative sur le secteur financier, je tiens à préciser ce que j'entends par ce terme. L'IA et l'IA générative sont des concepts certes liés, mais qui restent distincts. Si toute IA générative est de l'IA, toute IA n'est pas forcément générative.

L'IA est un vaste concept qui englobe tous les systèmes informatiques conçus pour imiter les fonctions cognitives humaines. Ces systèmes peuvent être classés en tant qu'IA étroite ou faible, c'est-à-dire conçue pour effectuer une tâche spécifique qui nécessitait auparavant l'intelligence humaine, ou bien en tant qu'IA générale (IGA) ou forte, c'est-à-dire capable d'appliquer une intelligence à un large éventail de tâches, comme le font les êtres humains.

Les systèmes d'IA qui occupent aujourd'hui le devant de la scène, comme ChatGPT, sont des exemples d'IA faible. Ils sont certes très avancés et capables de générer des textes similaires à ceux produits par des humains, mais ils se cantonnent à des tâches spécifiques et ne possèdent pas les capacités plus vastes d'une IGA. De nombreux chercheurs et experts s'accordent à dire que l'IGA complète reste hypothétique, car aucun système ne satisfait encore aux critères retenus. Toutefois, certains observateurs estiment qu'il ne s'agit désormais plus de savoir *si*, mais *quand*, cette IGA deviendra réalité (Fry, 2022).

L'IA générative fait référence à un système d'IA - qu'il soit faible ou fort - capable de générer un nouveau contenu ou de nouvelles données similaires, mais non identiques, aux données sur lesquelles il a été entraîné. Entraîner un algorithme consiste à lui présenter un vaste ensemble de données et à lui assigner une série de tâches à résoudre, en le guidant par le biais d'un système de récompense qui lui fournit un retour d'information sur ses performances. Ce processus permet à l'algorithme d'affiner sa compréhension des modèles et des critères de décision, ce qui améliore sa capacité à résoudre des tâches similaires à l'avenir. L'ensemble de données peut comprendre du texte, des images, de la musique, des documents vocaux et d'autres médias. Les modèles d'IA générative sont conçus pour comprendre les modèles et structures sous-jacents de ces données, ce qui leur permet de produire de nouveaux résultats originaux sur la base de ce qu'ils ont appris.

La caractéristique essentielle de l'IA générative est sa capacité à créer, qui offre un potentiel important en matière de création de contenu, de conception, voire de découverte de médicaments. Les capacités cognitives et créatives de l'IA générative suscitent continuellement de grands débats philosophiques (Millière & Buckner, 2024). L'IA générative a toutefois le potentiel non seulement de reproduire des tâches humaines, mais aussi d'accroître la créativité humaine en générant de nouvelles idées, de nouveaux concepts et de

nouvelles perspectives et de créer ainsi des applications innovantes qui dépassent le champ actuel des capacités d'analyse et de prise de décision de l'IA.

## Impact de l'IA générative dans le secteur financier

L'IA générative a déjà un impact significatif sur le secteur financier, avec des répercussions en termes de productivité, d'efficacité opérationnelle, de gestion des risques et de création de nouveaux modèles commerciaux. Selon un rapport de McKinsey, l'IA générative pourrait ajouter entre 200 et 340 milliards de dollars par an au secteur bancaire, soit l'équivalent de 9% à 15% de ses bénéfices d'exploitation (McKinsey & Company, 2023). Cette valeur découle principalement d'une augmentation de la productivité, les gains absolus les plus importants étant anticipés dans les secteurs des entreprises et de la vente au détail.

Grâce à l'IA générative, les institutions financières peuvent améliorer la gestion opérationnelle de leurs services et leur expérience client. Par exemple, McKinsey & Company (2023) rapporte que Goldman Sachs utilise un outil basé sur l'IA pour automatiser la génération de tests, qui nécessitait auparavant de grandes ressources en personnel. Citigroup a également eu recours à l'IA générative pour évaluer l'impact des nouvelles réglementations américaines en matière de fonds propres.

Maufe et Brown (2023) identifient cinq utilisations pratiques de l'IA générative dans le secteur des services financiers, qui augmentent toutes la productivité des employés. Il s'agit de la recherche et de la synthèse de documents financiers, des assistants virtuels améliorés, de la recherche sur les marchés financiers, du développement de logiciels et des recommandations financières personnalisées.

Toutefois, le Fonds monétaire international note que le déploiement de l'IA générative dans le secteur financier comporte plusieurs risques, tels que les biais intégrés, les lacunes en matière de protection de la vie privée, l'opacité du processus décisionnel, les problèmes de robustesse, les menaces de cybersécurité et les effets potentiels sur la stabilité financière au sens large. Le Fonds monétaire international met en garde qu'une dépendance excessive à l'égard de l'IA générative pourrait accroître le risque de contagion et créer de nouveaux risques systémiques dans le secteur financier, voir International Monetary Fund (2023).

Malgré ces préoccupations, le secteur financier explore activement les capacités de l'IA générative, les pressions concurrentielles encourageant l'adoption rapide d'applications qui promettent des gains d'efficacité, des économies de coûts et une meilleure gestion des risques.

La technologie de l'IA générative offre un éventail d'utilisations potentielles dans le domaine de la finance, notamment :

- **L'élaboration de prévisions et de budgets.** Une étude de KPMG a révélé que 83 % des participants utilisaient l'IA dans leur planification financière (KPMG, 2024), s'appuyant sur elle pour l'analyse prédictive, la création de scénarios et la fourniture d'informations budgétaires.
- **La création de rapports financiers et de présentations.** L'IA générative réduit considérablement le temps et les efforts nécessaires à la production de rapports financiers en alimentant automatiquement les modèles avec de nouvelles données et en tirant des enseignements des données financières et commerciales.
- **L'extraction d'informations sur le marché.** Grâce aux modèles de langage étendus (*Large Language Models – LLM*) de l'IA générative, les organisations peuvent puiser dans des ensembles de données publiques pour glaner des informations sur le marché et découvrir les tendances de leur clientèle par le biais d'analyses personnalisées qui peuvent être spécifiques à une région ou basées sur des personnes.
- **L'obtention d'informations stratégiques grâce aux données.** En utilisant l'IA générative pour analyser les données relatives à la gestion de la relation client (*Customer Relationship Management – CRM*) ou à la planification des ressources de l'entreprise (*Enterprise Resource Planning – ERP*), les entreprises peuvent obtenir des informations stratégiques précieuses pour éclairer leurs décisions financières, relever les défis en matière de tarification et de performance, ou améliorer d'autres opérations commerciales.
- **L'automatisation de la gestion des contrats.** L'IA générative rationalise la création des contrats, en se concentrant sur les termes non standard, en identifiant les clauses liées aux revenus et en documentant les traitements comptables.
- **L'amélioration de la détection des anomalies et la prévention des fraudes.** En identifiant les irrégularités dans les données et en surveillant les transactions en cours, l'IA générative offre un mécanisme robuste de détection des erreurs et de prévention de la fraude, ce qui permet de préserver l'intégrité financière d'une organisation.

- **La personnalisation des conseils d'investissement.** L'IA générative peut adapter les recommandations d'investissement au profil de risque et aux objectifs financiers d'un individu, améliorant ainsi l'expérience du client.
- **La détection des fraudes.** En analysant de vastes ensembles de données de transactions, l'IA générative peut identifier des schémas frauduleux, ce qui permet d'améliorer la sécurité financière.
- **L'analyse du marché et la gestion des risques.** À partir d'ensembles de données appropriés, l'IA générative peut identifier les tendances du marché et évaluer les risques potentiels, ce qui facilite les décisions d'investissement.

Cette grande variété d'applications dans le domaine financier soulève plusieurs points de discussion importants au sein de la profession. Ils concernent tout d'abord des considérations en matière d'éthique, en particulier la question de savoir comment maintenir la transparence. De nouvelles normes et de nouveaux protocoles sont nécessaires pour garantir que le contenu financier généré par l'IA ne perpétue pas des biais ou de la désinformation. Le deuxième point concerne l'impact de l'automatisation sur les employés du secteur financiers. Au fur et à mesure que l'IA supprime les rôles traditionnels, la main-d'œuvre du secteur financier doit acquérir de nouvelles compétences et être prête à s'adapter. Le troisième point se focalise sur le paysage réglementaire. Face aux avancées des technologies d'IA générative, les réglementations financières deviennent inadaptées ou obsolètes. Les cadres réglementaires doivent donc s'adapter pour à la fois encourager l'innovation et garantir la protection des consommateurs. Finalement, le développement continu de produits et de services innovants doit répondre aux demandes des marchés émergents et fournir aux institutions financières un avantage concurrentiel.

# Applications pratiques

A présent, examinons ensemble quelques applications pratiques de l'IA générative et explorons à quelles avancées elles contribuent dans le secteur financier. Ces exemples d'initiatives environnementales, sociales et de gouvernance (ESG) illustrent comment l'IA générative peut transformer un domaine spécifique de la finance et soulignent son potentiel à révolutionner l'ensemble du secteur.

## Étude de cas 1 : Identifier l'innovation verte

Dans un article récent, ma co-autrice et moi-même examinons comment l'IA générative joue un rôle crucial dans le décodage du langage quelque peu dense des brevets "verts" (Leippold et Yu, 2024). Comprendre les subtilités techniques de ces brevets peut être décourageant pour quelqu'un qui souhaite évaluer les initiatives d'une entreprise en matière de développement durable. Les modèles d'IA générative permettent de surmonter ce défi en transformant des descriptions complexes et une terminologie spécialisée en résumés concis. Le processus implique que l'IA "lise" systématiquement le brevet, interprète ses éléments cruciaux et les réarticule en termes simples. Cette application de l'IA générative accomplit de multiples fonctions essentielles pour le secteur financier.

Tout d'abord, en transformant un jargon technique en un langage vulgarisé, l'IA générative permet un accès plus large à des informations détaillées sur des projets d'innovation verte. Elle rend les informations contenues dans le brevet accessibles à un public diversifié, notamment aux analystes financiers, aux investisseurs et aux décideurs politiques, qui ne possèdent peut-être pas les connaissances techniques approfondies nécessaires pour les comprendre. Cette démocratisation de l'information est très importante, car elle permet à un plus grand nombre de parties prenantes de prendre des décisions en connaissance de cause.

Deuxièmement, les résumés fournis par l'IA améliorent l'efficacité de l'analyse. Les analystes financiers, qui travaillent dans des environnements où la rapidité et la précision sont primordiales, bénéficient grandement de la possibilité de comprendre rapidement l'importance d'un brevet sans se perdre dans les détails. Dans ce cas, l'IA permet non seulement de gagner du temps, mais aussi d'améliorer la qualité de l'analyse, ce qui garantit que les décisions financières sont basées sur une compréhension approfondie de l'impact potentiel et de la valeur des innovations d'une entreprise.

Troisièmement, les résumés générés grâce à l'IA facilitent l'analyse à l'échelle du marché. Lorsque les analystes financiers ont une compréhension claire des principales améliorations apportées par chaque brevet, ils peuvent plus aisément comparer les efforts en matière de développement durable de divers entreprises, secteurs

et industries. Cette capacité est très précieuse pour les investisseurs potentiels, car elle leur permet de s'assurer que leurs décisions sont fondées sur une compréhension complète de l'engagement des différentes entreprises en faveur de l'innovation verte.

Après avoir utilisé l'IA générative pour produire des résumés clairs et concis à partir de textes complexes de brevets verts, mes coauteurs et moi-même avons créé un modèle de langage basé sur BERT, baptisé *ClimateBERT*, afin d'affiner encore notre compréhension (Webersinke et al., 2022). BERT est l'acronyme anglais de *Bidirectional Encoder Representations from Transformers* (représentations d'encodeurs bidirectionnelles à partir de transformateurs). Développé par Google, BERT permet de mieux comprendre le contexte d'un mot en prenant en compte les mots qui le précèdent et le suivent dans la phrase. Nous avons entraîné ClimateBERT à reconnaître les nuances de la terminologie du développement durable, ce qui permet une analyse contextuelle approfondie des textes simplifiés produits par l'IA générative.

En analysant ces textes au fil du temps, ClimateBERT peut identifier des tendances dans le discours d'une entreprise sur le développement durable. Ce type d'analyse met en évidence les changements dans l'importance accordée par une entreprise à l'innovation verte. ClimateBERT fournit des signaux précoces de changements stratégiques qui pourraient affecter l'engagement à long terme d'une entreprise en faveur de la durabilité, ainsi que ses performances financières. En combinant l'IA générative et ClimateBERT, nous parvenons ainsi à une meilleure compréhension de la manière dont une entreprise s'engage dans l'innovation verte.

Enfin, pour comprendre les répercussions plus vastes de l'innovation verte, nous utilisons nos données pour créer des mesures qui montrent comment les initiatives de durabilité ont influencé les rendements anormaux cumulés, ajustés au marché, autour d'événements clés, tels que la victoire électorale de Trump en 2016, la victoire de Biden en 2020, le déclenchement de la guerre entre la Russie et l'Ukraine et l'annonce de l'*Inflation Reduction Act*. Nos travaux mettent à disposition des analystes financiers et des investisseurs un outil puissant pour mieux comprendre et pour réduire les risques liés à l'innovation verte et à la durabilité.

## Étude de cas n° 2 : analyse de la durabilité des entreprises

Mes coauteurs et moi-même avons développé un outil d'IA baptisé *CHATREPORT* (Ni et al., 2023), basé sur nos travaux antérieurs sur *CHATCLIMATE* (Vaghefi et al., 2023), afin de traiter les données denses et volumineuses contenues dans les rapports de durabilité des entreprises. Ces rapports détaillent l'impact environnemental, les structures de gouvernance et les initiatives sociales d'une

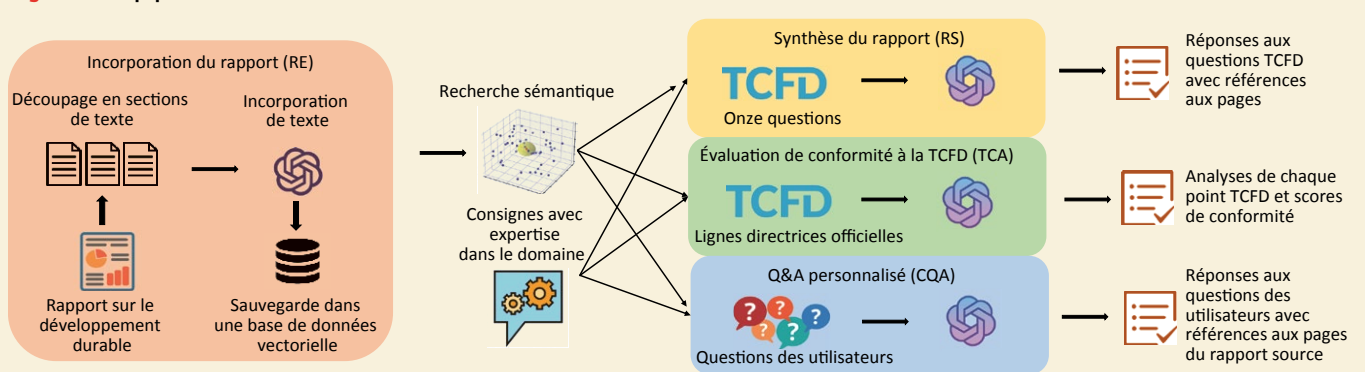
entreprise. Ils offrent un œil critique au travers duquel les investisseurs, les décideurs politiques et le grand public peuvent évaluer l'engagement d'une entreprise en matière de pratiques durables. Toutefois, en raison de leur complexité et du volume de données qu'ils contiennent, l'analyse manuelle de ces rapports n'est pas une mince affaire. Le défi est encore plus ardu lorsque l'analyse doit respecter des cadres réglementaires ou des lignes directrices spécifiques. Traditionnellement, seules quelques entités financières disposent des ressources nécessaires à une analyse approfondie et détaillée de ces rapports, ce qui constitue un obstacle important à la transparence. CHATREPORT démocratise le processus en permettant à un large éventail de parties prenantes de véritablement comprendre ces rapports et d'agir en conséquence.

Au cœur de CHATREPORT se trouve un pipeline sophistiqué, illustré dans la figure 1, qui transforme les données complexes des rapports de durabilité des entreprises en informations utiles pour la prise de décision. Un aspect essentiel de notre méthodologie consiste à intégrer l'expertise du domaine directement dans la boucle de développement de l'IA. Le processus commence par l'extraction du texte brut d'un rapport, la normalisation des formats de données et la suppression de toute information non pertinente. Pour comprendre et analyser le contenu, le pipeline utilise de grands modèles de langage (LLM) avancés qui sont affinés avec des connaissances spécifiques au domaine. Les utilisateurs et d'autres experts humains créent et affinent les consignes (*prompts*), en

veillant à ce que les résultats de l'IA soient techniquement exacts et contextuellement pertinents. L'outil identifie les thèmes clés, évalue la conformité avec des lignes directrices telles que celles de la *Task Force on Climate-related Financial Disclosures (TCFD)* et génère des informations dans un format d'utilisation plus convivial. Chaque étape du pipeline CHATREPORT est conçue pour garantir que les données soient traitées avec une grande précision et qu'elles respectent les dernières normes et pratiques en matière de durabilité. Le résultat est un outil robuste pour les parties prenantes qui ont besoin de prendre des décisions éclairées.

Comme nous le verrons plus loin, le risque d'"hallucination", c'est-à-dire le fait que l'IA produise des informations incorrectes ou trompeuses qui ne sont pas étayées par des données, constitue un problème courant dans les contenus générés par l'IA. CHATREPORT s'attaque à ce problème grâce à ses fonctions de traçabilité. Chaque élément de contenu généré par l'IA est directement traçable à des sections spécifiques des documents sources, ce qui permet à l'utilisateur de vérifier manuellement l'exactitude des informations. De surcroît, CHATREPORT est fortement axé sur les utilisateurs, offrant des fonctions qui permettent à ces derniers d'adapter l'analyse à leurs besoins. Grâce à un module de questions-réponses dynamique, les utilisateurs peuvent interroger l'IA sur des aspects particuliers d'un rapport. L'IA récupère et traite alors les données pertinentes, ce qui permet de fournir des informations personnalisées.

**Figure 1: Le pipeline CHATREPORT**

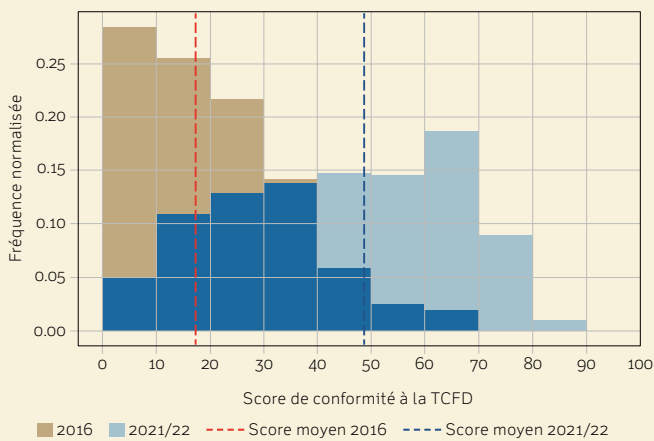


Note: Le pipeline de CHATREPORT automatise l'analyse des rapports de durabilité des entreprises grâce à quatre modules principaux. Le module *Report Embedding (RE)* divise le rapport en morceaux de texte, qu'il intègre dans un espace vectoriel pour mener une recherche sémantique basée sur les recommandations de la TCFD. Le module *Report Summarization (RS)* récupère les sections pertinentes du rapport et demande à un modèle de langage étendu (LLM) de résumer les informations en fonction des onze aspects recommandés par la TCFD. Le module *TCFD Conformity Assessment (TCA)* évalue l'adhésion du rapport aux lignes directrices de la TCFD, en générant des scores de conformité et des paragraphes d'analyse. Le module *Customized Question Answering (CQA)* permet aux utilisateurs de poser des questions spécifiques, le LLM fournissant des réponses se référant aux sections correspondantes du rapport. Ce pipeline complet améliore la transparence et l'efficacité de l'analyse des rapports de développement durable.

Source: Ni et al. (2023)



**Figure 2: Graphique de densité des notes de conformité à la TCFD**



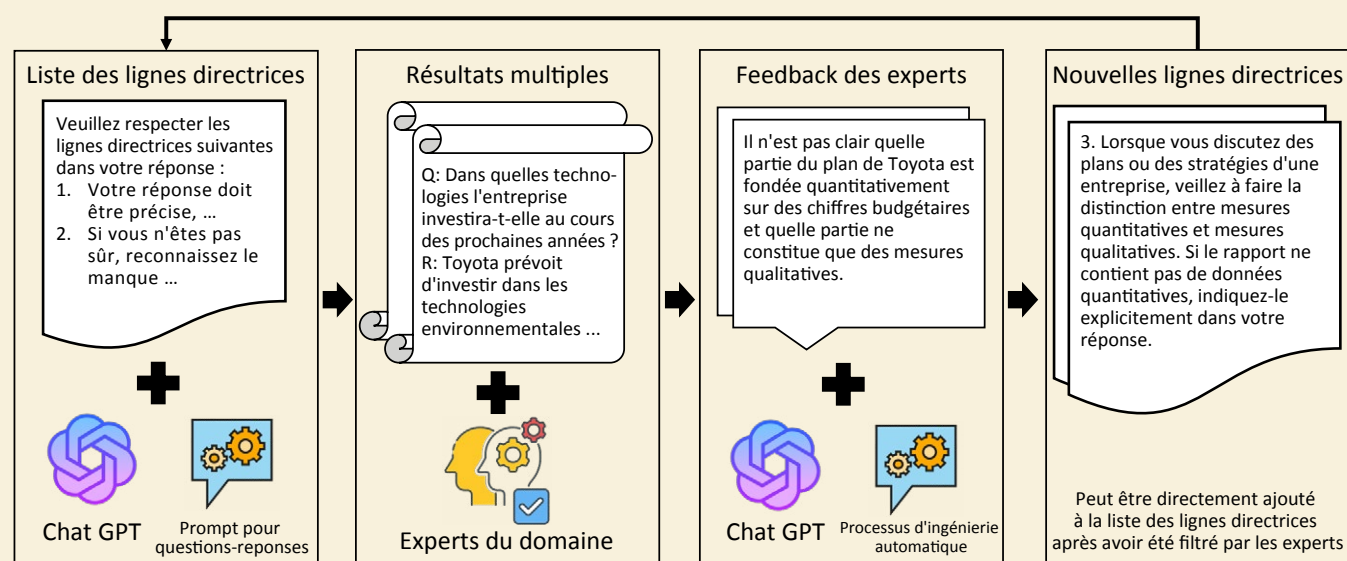
Note: Cette figure illustre la distribution des scores de conformité à la TCFD pour un échantillon de rapports de développement durable d'entreprises cotées à la Bourse de New York, pour les années 2016 et 2021-2022. Les scores *TCFD Conformity Assessment* sont calculés en fonction du degré d'adhésion des rapports aux lignes directrices de la TCFD. Le graphique de densité montre une amélioration notable des scores de conformité moyens entre les années 2016 et 2021-2022, ce qui indique que les entreprises suivent de plus en plus les recommandations de la TCFD concernant les informations qu'elles divulguent en matière de développement durable.

Source: Ni et al. (2023)

Dans notre étude, par exemple, nous avons évalué plus de 10'000 rapports d'entreprise sur le développement durable pour vérifier leur conformité avec les exigences de la TCFD en matière de reporting. La figure 2 illustre visuellement les scores pour deux périodes de reporting: avant la TCFD, en brun, et quelques années après la TCFD, en bleu clair. Nos données indiquent que le niveau de conformité aux lignes directrices de la TCFD a considérablement augmenté entre ces deux périodes. Avant l'inauguration officielle de la TCFD, en 2016, le score de conformité moyen était de 17 %. En 2021-2022, il atteignait 49 %. Cette tendance met en évidence l'importance croissante accordée à la transparence et à l'exhaustivité des informations financières liées au climat. Dans des analyses plus approfondies, nous avons identifié des tendances et des valeurs aberrantes, ce qui permet d'alimenter le débat sur l'efficacité des pratiques actuelles et sur les domaines potentiels d'amélioration.



**Figure 3: Pipeline d'ingénierie de prompt automatique**



Note : CHATREPORT intègre les commentaires des utilisateurs et d'autres experts du domaine afin d'affiner et d'améliorer les prompts utilisés par le LLM. Dans un premier temps, des experts du domaine fournissent un retour d'information sur les réponses spécifiques générées par le LLM. Ce feedback est ensuite transformé en lignes directrices générales par le biais d'un processus automatisé d'ingénierie de prompts. L'IA intègre ces nouvelles lignes directrices dans ses modèles, ce qui permet d'améliorer la qualité et la précision de ses futurs résultats. Ce processus itératif permet une analyse plus précise, plus détaillée et plus critique des rapports de durabilité.

Source: Ni et al. (2023)

Enfin, comme le montre la figure 3, CHATREPORT est conçu pour apprendre et s'adapter au fil du temps. Le système intègre en permanence les commentaires des utilisateurs et d'autres experts humains pour affiner ses modèles analytiques et ses prompts. Ce processus itératif garantit que le système s'améliore constamment en termes de précision, de fiabilité et de pertinence. Dans un souci de transparence, mes co-auteurs et moi-même avons développé CHATREPORT en tant que projet *open-source*. Il est accessible sur <https://reports.chatclimate.ai/>. En mettant à la disposition du public nos méthodologies, nos ensembles de données et nos outils, nous encourageons la collaboration et invitons les développeurs et les chercheurs du monde entier à contribuer à l'avancement de l'analyse financière durable.

La puissance de CHATREPORT réside dans sa capacité à accélérer considérablement le processus d'analyse. Grâce à cet outil d'IA, des informations pertinentes peuvent être extraites de milliers de rapports d'entreprise sur le développement durable (jusqu'à 10'000) en quelques heures seulement. CHATREPORT condense rapidement de vastes quantités de données en résumés concis, ce qui permet une prise de décision plus rapide et davantage fondée sur les données, à une échelle jusqu'alors inaccessible.

### Un aperçu de l'avenir (très) proche

L'IA générative pourrait bientôt modifier radicalement la manière dont les investissements et les décisions financières sont analysés. J'ai récemment testé une approche de *fact-checking* automatisé qui rappelle la méthode socratique. Dans la Grèce antique, Socrate interrogeait sans relâche ses interlocuteurs pour mettre en évidence les contradictions dans leurs pensées, encourageant ainsi une exploration plus profonde des questions sous-jacentes. Cet héritage philosophique se retrouve dans l'approche de l'outil *Mediator-Advocate* employé par Leippold et al. (2024) pour vérifier les affirmations sur les questions climatiques.

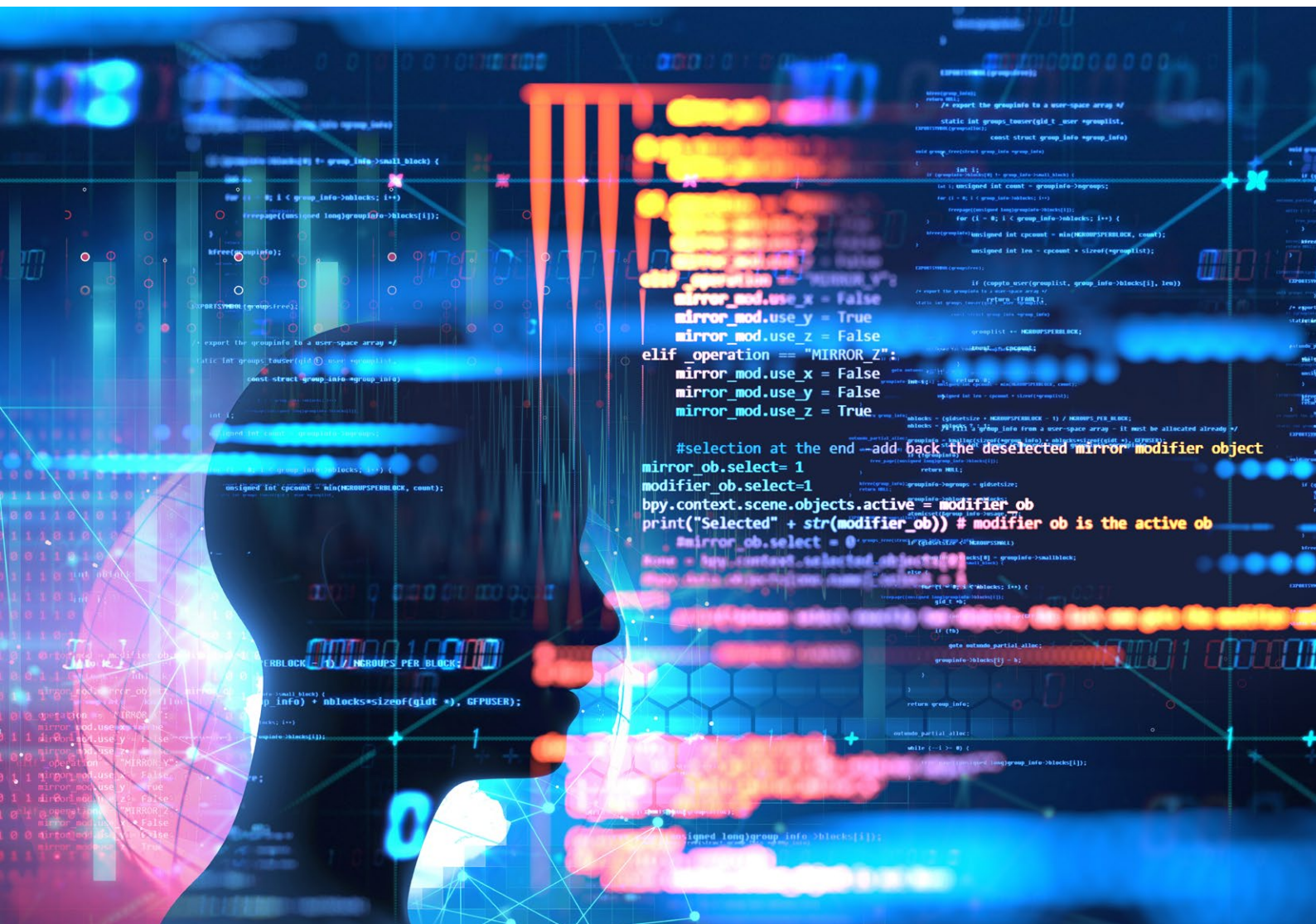
Si nous appliquions cet outil au domaine de la finance, nous aurions plusieurs LLM, appelés "analystes IA", chacun analysant une facette spécifique des données financières (telles que les tendances du marché, les impacts réglementaires, les indicateurs économiques ou les mesures de performance des entreprises) et proposant des stratégies d'investissement potentielles. Ces arguments seraient ensuite intégrés par un "responsable des investissements IA". Pour obtenir une évaluation cohérente, ce responsable poserait des questions complémentaires aux analystes IA, de la même

manière que Socrate guidait ses élèves vers une compréhension plus profonde grâce à la dialectique. Ce processus garantit que de nombreux points de vue différents soient pris en compte, qu'ils s'opposent entre eux et se complètent mutuellement, pour finalement aboutir à une conclusion plus solide.

Un tel outil améliorerait la prise de décision financière en tirant parti de la vitesse de calcul de l'IA, ainsi qu'en réduisant les biais en faveur d'un type d'analyse financière plutôt qu'un autre. En outre, les marchés financiers devenant de plus en plus complexes, l'évolutivité des systèmes d'IA permettrait aux décideurs de traiter efficacement la hausse croissante des données.

Une caractéristique essentielle d'un tel système serait sa capacité à suivre et à documenter la progression des arguments entre les analystes IA, en fournissant une vérification rétrospective sur la manière dont ils sont parvenus à leurs décisions. Cette traçabilité est essentielle pour renforcer la confiance entre les parties prenantes et pour répondre aux exigences réglementaires du secteur financier, qui nécessitent une documentation complète des processus de prise de décision, en particulier dans des domaines tels que l'évaluation des risques et la mise en conformité.

L'intégration des systèmes d'IA dans la finance promet une évolution vers une prise de décision plus dynamique et en temps réel. Cependant ce potentiel s'accompagne de nombreux défis théoriques et pratiques, tels que la garantie de l'exactitude des données, le maintien de la crédibilité des sources et l'atténuation des risques de biais algorithmiques.



# Limites actuelles de l'IA générative

Malgré son potentiel, la technologie de l'IA générative souffre d'importantes limitations qui nécessitent une attention urgente. Avant qu'elle ne puisse révolutionner le secteur financier, des problèmes tels que l'hallucination, le manque de vérifiabilité et la vulnérabilité aux attaques adverses devront être résolus.

## Hallucination et imprécision

Dans Schimanski et al. (2024), mes co-auteurs et moi-même analysons l'exactitude et la vérifiabilité des réponses générées par les LLM. Dans les tâches complexes de réponse aux questions et de génération de contenu, le phénomène connu sous le nom d'"hallucination" est courant. Dans le domaine de l'IA, l'hallucination désigne le fait que le modèle génère des informations qui sont certes plausibles, mais n'ont rien à voir avec les sources réelles et sont totalement infondées. Ce problème est crucial dans des secteurs tels que la finance, la santé et le droit, où l'exactitude et la fiabilité des informations sont primordiales. Notre étude met en évidence la façon dont les LLM attribuent souvent incorrectement des sources ou fabriquent du contenu erroné, ce qui met sérieusement à mal la confiance des utilisateurs dans ces systèmes automatisés.

Pour s'attaquer à ce problème, nous proposons d'améliorer le processus d'extraction de texte à partir des sources pertinentes et d'améliorer l'attribution du contenu généré par l'IA. Ces correctifs techniques ne constituent toutefois qu'une partie de la solution. Les répercussions d'une inexactitude non corrigée sont vastes, avec des conséquences significatives possibles sur les décisions financières, les soins aux patients et les procédures judiciaires. Il est urgent de mener des travaux de recherche pour affiner ces modèles et élaborer un cadre éthique solide pour leur utilisation.

## Manque de vérifiabilité

Les moteurs de recherche génératifs, y compris des plateformes bien connues telles que *Bing Chat*, *NeevaAI*, *perplexity.ai* et *YouChat*, sont conçus pour répondre aux requêtes des utilisateurs en générant des réponses complètes avec des citations en ligne. Cependant, Liu et al. (2023) constatent que seulement 52% des phrases produites par ces moteurs de recherche sont correctement étayées par des citations. Ce chiffre est alarmant, compte tenu de l'importance d'informations fiables dans notre ère numérique. De plus, l'étude montre que seulement 75% des citations appuient réellement les déclarations qu'elles sont censées valider. Cet écart entre apparence et réalité induit en erreur les utilisateurs, qui se fient à des informations apparemment crédibles, mais fondamentalement non étayées ou incorrectement citées.

Cette recherche met en lumière un aspect crucial du développement de l'IA, soit sa fiabilité. Malgré les résultats sophistiqués de ces moteurs de recherche génératifs, la prévalence d'affirmations non étayées et d'inexactitudes dans les citations nuit considérablement à leur fiabilité. La formation de l'IA doit être affinée et des stratégies développées pour améliorer la vérifiabilité des informations fournies par ces systèmes.

## Vulnérabilité aux attaques adverses

L'IA générative est également sensible aux attaques adverses qui manipulent les systèmes d'IA afin d'intentionnellement mal interpréter ou déformer les données. Cette faiblesse est particulièrement inquiétante dans certains types d'analyses financières, où de telles manipulations peuvent gravement fausser le résultat de l'analyse. Dans Leippold (2023), par exemple, j'ai étudié les attaques de type "*sentiment spin*", dans lesquelles un modèle sophistiqué comme GPT-3 peut être utilisé pour modifier la tonalité de textes financiers, en modifiant une tonalité négative d'une série de phrases, par exemple, en un sentiment neutre ou positif. De telles attaques posent de sérieux risques pour la fiabilité des systèmes automatisés.

Hu et al. (2024) détaillent diverses autres stratégies adverses qui exploitent les vulnérabilités des moteurs de recherche génératifs. Leur recherche montre comment les moteurs de recherche génératifs, malgré leurs capacités avancées, peuvent être induits en erreur par des requêtes spécifiquement conçues pour ce faire. Cette vulnérabilité met en évidence le besoin urgent d'améliorer les mesures de sécurité de l'IA.

Les implications des attaques adverses ne se limitent pas à la finance. Les systèmes d'IA peuvent également être manipulés pour produire des résultats biaisés ou trompeurs dans d'autres secteurs tels que la santé, la politique et la pharma. La réglementation de l'IA générative devrait se concentrer sur la création de cadres qui garantissent la transparence, la responsabilité et l'équité. Au fur et à mesure que l'utilisation de l'IA se généralise, les régulateurs doivent mettre en œuvre des lignes directrices strictes afin de

promouvoir des pratiques éthiques et de prévenir les abus. De telles réglementations protégeront les utilisateurs des effets des attaques adverses et favoriseront la confiance dans les systèmes d'IA, ouvrant ainsi la voie à une acceptation et une intégration plus large de l'IA dans la société.

Bien que certains cadres existent, comme la "loi européenne sur l'intelligence artificielle", il reste encore beaucoup à faire pour élaborer des réglementations complètes et efficaces. La formulation de ces réglementations nécessite une approche nuancée qui tienne compte du rythme rapide de l'évolution technologique et de la diversité et de l'expansion des applications de l'IA. Il est essentiel de maintenir un dialogue permanent entre les universitaires, les dirigeants d'entreprise - en particulier du secteur technologique - et les décideurs politiques.



# Conclusion

L'IA générative s'accompagne à la fois d'opportunités sans précédent et de vastes défis. Dans cette Public Discussion Note, j'ai mis en avant son potentiel de transformation pour le secteur financier, qu'il s'agisse d'améliorer la prise de décision et l'évaluation des risques ou de révolutionner les pratiques de durabilité des entreprises. Cependant, comme nous l'avons évoqué avec les problèmes d'hallucination, de vérifiabilité et de vulnérabilité aux attaques adverses, le chemin vers l'intégration complète de l'IA dans le domaine de la finance est semé d'embûches qui pourraient miner la confiance dans les systèmes d'IA et nuire à leur fiabilité.

Les investissements substantiels déjà réalisés dans l'IA soulignent l'importance des enjeux. Selon Goldman Sachs (2024), si le potentiel de gains de productivité et d'efficacité opérationnelle est évident, les avis divergent quant à l'impact économique global de l'IA. Des experts comme Daron Acemoglu du *Massachusetts Institute of Technology*, entre autres, expriment leurs doutes. Ils craignent que

la technologie de l'IA ne justifie pas ses coûts élevés, qui pourraient ne pas diminuer comme prévu, voir Acemoglu (2023). A l'inverse, de nombreux observateurs plus optimistes estiment que l'IA finira par produire des rendements et des avantages économiques substantiels, et que nous n'avons pas encore vu émerger ses applications les plus transformatrices.

En tant que chercheur dans ce domaine en pleine évolution, je suis conscient à la fois du potentiel de l'IA et des défis futurs liés à son application dans la finance. L'avenir de l'IA générative dans la finance n'est pas prédéterminé, mais sera façonné par nos actions. Nous avons une occasion unique d'exploiter l'IA pour stimuler l'innovation, renforcer la durabilité et améliorer l'efficacité au sein du secteur financier. Toutefois, atteindre ces objectifs nécessite une mise en œuvre prudente et réfléchie afin d'atténuer les risques posés par une exubérance incontrôlée.



# Références

- Acemoglu, D. (2023). Harms of AI. In *The Oxford Handbook of AI Governance* (pp. 660–706). Oxford University Press eBooks. <https://doi.org/10.1093/oxfordhb/9780197579329.013.65>
- Cookson, C. (2023, November 14). AI outperforms conventional weather forecasting methods for first time. *Financial Times*. <https://www.ft.com/content/ca5d655f-d684-4dec-8daa-1c58b0674be1>
- Fry, H. [Host] (2022). *DeepMind: The podcast – The road to AGI* [Audio podcast]. Google. <https://deepmind.google/discover/the-podcast/the-road-to-agi/>
- Golman Sachs (2024). *Gen AI: Too much spend, too little benefit?* <https://www.goldmansachs.com/intelligence/pages/gi-research/gen-ai-too-much-spend-too-little-benefit/report.pdf>
- Hu, X., Li, X., Chen, J., Li, Y., Li, Y., Li, X., Wang, Y., Liu, Q., Wen, L., Yu, P. S., & Guo, Z. (2024). Evaluating robustness of generative search engine on adversarial factual questions. arXiv.org. <https://arxiv.org/abs/2403.12077>
- International Monetary Fund (2023). *Generative artificial intelligence in finance: Risk considerations*. <https://www.imf.org/-/media/Files/Publications/FTN063/2023/English/FTNEA2023006.ashx>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. a. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- KPMG (2024). *Why finance should lead the adoption of generative AI*. <https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2024/generative-ai-finance.pdf>
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science* 382(6677), 1416–1421. <https://doi.org/10.1126/science.adi2336>
- Leippold, M. (2023). Sentiment spin: Attacking financial sentiment with GPT-3. *Finance Research Letters* 55(B), 103957. <https://doi.org/10.1016/j.frl.2023.103957>
- Leippold, M., Vaghefi, S. A., Stambach, D., Muccione, V., Bingler, J., Ni, J., Colesanti-Senni, C., Wekhof, T., Schimanski, T., Gostlow, G., Yu, T., Luterbacher, J., & Huggel, C. (2024). *Automated fact-checking of climate change claims with large language models*. arXiv.org. <https://arxiv.org/abs/2401.12566>
- Leippold, M., & Yu, T. (2024). *Are green innovations priced? Evidence beyond patents*. Swiss Finance Institute Research Paper. <http://dx.doi.org/10.2139/ssrn.4391444>
- Liu, N. F., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. arXiv.org. <https://arxiv.org/abs/2304.09848>
- Maufe, Z., & Brown, T. (2023, October 3). *Five generative AI use cases for the financial services industry*. Google Cloud. <https://cloud.google.com/blog/topics/financial-services/five-generative-ai-use-cases-financial-services-industry>
- McKinsey & Company (2023). *Capturing the full value of generative AI in banking*. <https://www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking#/>
- Millière, R., & Buckner, C. (2024). *A philosophical introduction to language models – Part I: Continuity with classic debates*. arXiv.org. <https://arxiv.org/abs/2401.03910>
- Munger, C. (2023, May 6). Answer [Q&A Session]. Berkshire Hathaway Annual Meeting, Omaha, NE, United States.
- Ni, J., Bingler, J., Colesanti-Senni, C., Kraus, M., Gostlow, G., Schimanski, T., Stambach, D., Vaghefi, S. A., Wang, Q., Webersinke, N., Wekhof, T., Yu, T., & Leippold, M. (2023). *CHATREPORT: Democratizing sustainability disclosure analysis through LLM-based tools* [Conference proceedings]. 2023 Conference on Empirical Methods in Natural Language Processing, Singapore. <https://doi.org/10.18653/v1/2023.emnlp-demo.3>
- Schimanski, T., Ni, J., Kraus, M., Ash, E., & Leippold, M. (2024). Towards faithful and robust LLM specialists for evidence-based question-answering. arXiv.org. <https://arxiv.org/abs/2402.08277>
- Vaghefi, S. A., Stambach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., Allen, S., Colesanti-Senni, C., Wekhof, T., Schimanski, T., Gostlow, G., Yu, T., Wang, Q., Webersinke, N., Huggel, C., & Leippold, M. (2023). ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment* 4(1). <https://doi.org/10.1038/s43247-023-01084-x>
- Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). *ClimateBert: A pretrained language model for climate-related text*. arXiv.org. <https://doi.org/10.48550/arXiv.2110.12010>

## Swiss Finance Institute

Avec le soutien de ses fondateurs – le secteur bancaire suisse, la Confédération et les principales universités suisses – le Swiss Finance Institute (SFI) assure une promotion active d'une recherche et d'un enseignement de classe mondiale dans les domaines de la banque et de la finance en Suisse. En combinant excellence académique et expérience pratique, le SFI contribue au renforcement de la place financière suisse.

### Editeur et contact

Dr. Cyril Pasche  
Senior Director Publications and Topic Development  
+41 22 379 88 25  
cyril.pasche@sfi.ch

swiss:finance:institute

Walchestr. 9, CH-8006 Zurich, T +41 44 254 30 80  
info@sfi.ch, www.sfi.ch

